Collaborative genome-wide association analysis supports a role for *ANK3* and *CACNA1C* in bipolar disorder

SUPPLEMENTARY INFORMATION

Outline

1. Supplementary Methods

Individual datasets

This study combined individual genotyping data from two published GWAS of BD, the WTCCC[1] and STEP-UCL[2] studies, and a previously unpublished dataset (ED-DUB-STEP2). All analyses were performed with the PLINK package[3]. We used the WTCCC Affymetrix 500K genotyping data (2007-02-05 CHIAMO data freeze calls, confidence score >0.9) available on the WTCCC website for the BD cases (n = 1,998) and the two control groups (n = 3,004). We excluded 196 individuals and 30,956 SNPs that failed the WTCCC quality control (QC) filters. We also excluded WTCCC individuals with high levels of average genome-wide identity-by-state (IBS) sharing with other WTCCC (n = 4), STEP-UCL (n = 31) or ED-DUB-STEP2 (n = 7) individuals, and SNPs that were reported by the WTCCC to have bad clustering (n = 578) or had a minor allele frequency (MAF) < 0.01 in the combined sample of BD cases and shared controls (n = 63,290). As a result, for the present analyses, the WTCCC panel consisted of 1,829 cases and 2,935 controls genotyped for 405,774 SNPs. The original STEP-UCL dataset comprised 1,461 BD cases and 2,008 controls, also genotyped in the Affymetrix 500K array. In total, 372,193 SNPs passed QC and had a MAF > 0.01 as described previously[2]. We excluded one case and one control that showed high genome-wide IBS sharing with a WTCCC or ED-DUB-STEP2 individual. The ED-DUB-STEP2 dataset consisted of 1,098 BD cases and 1,267 controls genotyped using the Affymetrix 500K, 5.0 or 6.0 arrays. After QC, there remained 331,786 SNPs for analysis. A more detailed description of sample ascertainment and QC for this study is available in the "Quality control protocol for ED-DUB-STEP2" section below.

Main phenotype definition

Bipolar disorder (BD) refers to an episodic recurrent pathological disturbance in mood, ranging from extreme elation or mania to severe depression, and usually accompanied by disturbances in thinking and behavior; psychotic symptoms (delusions and hallucinations) often occur. All BD cases had suffered one or more episodes of pathologically elevated mood, a criterion that captures the clinical spectrum of bipolar mood variation that shows familial aggregation[4]. The distribution of cases amongst DSM-IV[5] diagnostic categories is shown in Supplementary Table 3 online. We have analyzed the broad bipolar disorder category that includes BP1, BP2 and schizoaffective disorder, as twin and family studies suggest that these are likely to have a large set of overlapping genetic risk factors[6-10]. Results remained largely unchanged when considering BP1 cases only (not shown).

Subphenotype definitions

Age-at-onset was defined as the age at which the first impairment occurred, coded for the present analysis as >18 or ≤ 18 years of age.

For the WTCCC, non-psychotic cases were those with either (a) no lifetime definite psychotic symptoms (delusions or hallucinations) or (b) with only a single known brief psychotic event which was not a prominent part of any episode of illness. Cases with multiple definite psychotic symptoms were classified as psychotic.

In STEP-BD, psychotic cases were those who had any of the following: a history of psychotic affective disorder based on the ADE diagnostic interview, met criteria for a psychotic disorder based on the MINI diagnostic interview, hallucinations or delusions based on the Young Mania

Rating Scale at baseline or during follow-up, or clinician ratings of moderate to severe hallucinations, delusions, paranoid ideation or ideas of reference on the Clinical Monitoring Form at baseline or follow-up. Non-psychotic cases were those who were negative for psychosis based on the above measures.

For the UCL sample, bipolar cases were defined as psychotic if they had grandiose delusions, hallucinations, paranoid delusions or ideas of reference during a mania as shown by the SADS-L, OPCRIT data and review of medical case notes. Non-psychotic cases were those who were negative for psychosis based on the above measures.

For the Edinburgh sample, patients with bipolar disorder were defined as psychotic by a trained psychiatrist if delusions or hallucinations during an episode of mania were recorded as present in the SADS-L interview or described in the case records. Non-psychotic cases were those who were negative for psychosis based on the above measures.

Combined dataset

The three individual datasets (WTCCC, STEP-UCL and ED-DUB-STEP2) were merged, retaining only overlapping SNPs (n = 326,201). Of these, we excluded SNPs that had a Hardy-Weinberg equilibrium $P < 10^{-6}$ in controls (n = 433) or a significant ($P < 10^{-6}$) allele frequency difference between the three control groups and that were likely to represent technical artifacts rather than true population differences (n = 78). The combined dataset consisted of 10,596 individuals (4,387 cases, 6,209 controls) and 325,690 genotyped SNPs. The most associated SNPs ($P < 10^{-6}$) were also tested for association after expanding the WTCCC control group to include the six additional WTCCC disease groups ("expanded reference group" analysis), as this strategy can improve

power. This increased the total number of controls from 6,209 to 17,582. Throughout, genomic positions correspond to the NCBI build-36 of the human genome and SNP alleles are expressed in the forward (+) strand of the same build.

Imputation analysis

We developed the following approach to impute genotype data for SNPs not directly genotyped in a GWAS dataset but present on a reference panel, such as the HapMap. This approach, which is implemented in version v1.01 of PLINK, is an extension of multi-marker tagging. In the text below, an "observed" SNP refers to one that was genotyped in both the reference and the GWAS sample, whereas an "imputed" SNP refers to one that only appears in the reference panel.

We first create a single fileset with the reference panel merged in with the GWAS datset, in this case the WTCCC, STEP-UCL and ED-DUB-STEP2 combined dataset. The reference panel used for the present analysis consisted of 2,543,285 polymorphic autosomal SNPs genotyped on the 60 HapMap CEU founders and available for download from www.hapmap.org or, in PLINK format, from http://pngu.mgh.harvard.edu/~purcell/plink/res.shtml. The resulting dataset contained 325,690 SNPs genotyped and 2,233,448 SNPs with missing data for the GWAS individuals.

For each SNP with missing data for the GWAS individuals ("reference" SNP), we selected a set of neighboring observed SNPs ("proxy" SNPs) using the reference panel information, and then phased these proxy SNPs using a standard EM algorithm in both reference panel and GWAS sample jointly. By grouping subsequently haplotypes based on the allele at the reference position, genotype data for the reference SNP can then be inferred, either as discrete "hard" genotype calls,

or fractional allelic dosage. For the fractional allelic dosage, a value of 1 indicates two copies of the minor allele, whereas a value of 0 indicates zero copies of the minor allele.

For the present analysis, we selected up to 5 SNPs either side of the reference SNP, from a search of 15 SNPs either side of the reference, at most 250 kb away. All potential proxies were examined one at a time in order of strongest to weakest LD with the reference. A proxy must be above r^2 0.05 with the reference to be added; additionally, a proxy must not have an r^2 greater than 0.5 with any proxy already selected. An exception is that we always attempt to add at least two proxies even if the best proxy does not meet the $r^2 > 0.05$ criterion, to ensure we always find a 2-SNP haplotype that might tag the reference SNP, even if no single SNP does. In addition, proxies were only selected if the MAF ≥ 0.01 and genotyping failure rate < 0.05. For the present analysis, we used this approach to impute unmeasured genotypes and test these for association on-the-fly.

We obtain concordance rate information for all observed SNPs via a "leave-one-out" procedure. Specifically, each observed SNP is dropped from the GWAS dataset, one at a time, and then re-imputed given the surrounding proxy SNPs. An information score is generated for each imputed SNP that reflects how confidently genotypes were inferred. This is calculated as the ratio of the empirical variance of the imputed allele dosage scores for all individuals against the theoretical variance given the imputed minor allele frequency and assuming Hardy-Weinberg equilibrium. Values below 0.80 are taken to be indicative of poor quality imputation.

Using this approach, we inferred genotypes for 2,233,448 autosomal HapMap SNPs that were not present in the combined dataset. A total of 1,444,258 SNPs were imputed with high confidence (information score ≥ 0.80) and tested together with genotyped SNPs in the primary association

analysis. The average concordance between inferred and observed genotypes for SNPs imputed with high confidence was 0.987 (Supplementary Table 4 online). Similar results were obtained with MACH1[11] (data not shown).

Association analysis

The primary analysis was a logistic regression of disease state on single SNP allele dosage, either directly genotyped or imputed (in which case dosage could take any value between 0 and 1), including covariates to account for study effects (coded as six dummy variables) and the first two quantitative indices of ancestry based on a multi-dimensional scaling analysis. No clinical variables were included as covariates. For the expanded control analysis, which was only performed for the three regions of strongest association, indices of ancestry were not used. Including imputed SNPs, we tested 1,769,948 variants. The genomic inflation factor (λ) was calculated as the ratio between the observed and expected median chi-square statistics and was used to assess, but not correct for, the degree of inflation of the test statistic as a result of unaccounted sources of bias. As λ increases with sample size, we re-scaled λ for a study of 1,000 cases and 1,000 controls[12-14] (λ_{1000}). Throughout, odds ratios (OR) give the increased or decreased odds for the minor allele over the major allele and were obtained from the logistic regression analysis. The Breslow-Day test was used to compare OR between studies.

Treatment response

To investigate whether the regions of strongest association influenced response to treatment, we examined 447 cases drawn from the STEP-BD cohort who were treated exclusively with lithium (n=274) or lamotrigine (n=173). Individuals were followed prospectively for up to 2 years, with visits as often as clinically indicated, but no less than once every 4 months. For survival analysis,

time-to-event was defined as number of days following first euthymic visit (i.e., baseline) to the first recurrence of a DSM-IV-defined major depressive, hypomanic, manic or mixed episode, with results censored after recurrence, dropout, completion of 2 years of follow-up, or presence of a gap between visits of greater than 120 days[15]. Time to recurrence was assessed using Cox regression, which included the two indices of ancestry and a term for DSM-IV diagnosis.

Two-locus analysis

We developed the following approach for the detection of SNP × SNP pairwise interactions in large-scale case-control association studies. We follow the procedure for constructing an allelic test of a single locus, twice collapsing three genotype categories into two allele categories. Specifically, we count the 4*N* independent alleles observed at two loci in a sample of *N* individuals into a 2x2 table, following the logic below, so the allele (not the individual or haplotype) is the unit of analysis.

Table S1

0

	BB	**Bb**	**bb**
AA	a	b	c
Aa	d	e	f
aa	g	h	i

→

(b)

	B	**b**
A	4^a+2b+2d+e	4c+2b+2f+e
a	4g+2h+2d+e	4i+2h+2f+e

Based on the 2x2 Table S1(b), the odds ratio and standard error between loci A and B are calculated in the standard manner. When cases and controls are present, the above procedure is

performed separately in cases and controls, and the test for epistasis is the difference of the two odds ratios:

$$\mathbf{Z} = \mathbf{abs}\left(\frac{\log\left(\Phi_{\mathbf{Cases}}\right) - \log\left(\Phi_{\mathbf{Controls}}\right)}{\sqrt{\mathbf{SE}\left(\log\left(\Phi_{\mathbf{Cases}}\right)\right) + \mathbf{SE}\left(\log\left(\Phi_{\mathbf{Controls}}\right)\right)}} \right)$$

where $\Phi = \left(\mathbf{c}_{11} \cdot \mathbf{c}_{22}\right) / \left(\mathbf{c}_{12} \cdot \mathbf{c}_{21}\right)$, $\mathbf{SE}(\log(\Phi)) = 1/\mathbf{c}_{11} + 1/\mathbf{c}_{12} + 1/\mathbf{c}_{21} + 1/\mathbf{c}_{22}$, and c_{11}, c_{12}, c_{21} and c_{22} represent the four cells of Table S1(b). This test follows a standard normal distribution under the multiplicative model of no interaction.

We observe appropriate type I error rates in simulation (Table S2) and equivalent power to the logistic regression test. The correlation with a logistic regression analysis is very high (r = 0.995, based on –log10 P-value). The power to detect a large interaction effect (GRR = 2) and no marginal single SNP effects was 0.74 ($\alpha = 1.2 \times 10^{-12}$, disease prevalence = 0.01, MAF = 0.1 for both loci). Power for other two-locus models can be estimated using the power calculator available through http://pngu.mgh.harvard.edu/~purcell/gpc/.

***Table S2**:* Type-I error of the case-control epistasis test. We considered three models that included no interaction between two unlinked SNPs and no marginal SNP effect (model 1) or a strong effect for one (model 2) or both SNPs (model 3). Type-I error is based on the analysis of 100,000 simulated datasets (disease prevalence = 0.01, minor allele frequency = 0.1 for both loci).

Model	Marginal SNP effects (Odds Ratio)		Nominal alpha	
	SNP 1	SNP2	0.05	0.0005
1	1.0	1.0	0.04750	0.00030
2	1.0	1.4	0.04941	0.00030
3	1.4	1.4	0.04817	0.00048

Our procedure assumes Hardy-Weinberg and linkage equilibrium for the two SNPs hold in the population. However, simulation studies have shown the case/control test to be very robust to deviations from the linkage equilibrium assumption, whereas a case-only test is not (data not shown). Analogous to adopting an allelic single locus test, we also assume an allelic mode of gene action where any interaction term represents an allele-by-allele effect, not a genotype-by-genotype effect.

We use this procedure as a computationally efficient screen of the more than 40 billion pairwise comparisons and confirm all results with $P < 5\times10^{-10}$ by logistic regression. Based on the 290,488 genotyped autosomal SNPs with a MAF ≥ 0.05, we applied this approach to all pairs of SNPs for a total of 42,191,493,828 tests.

Quality control (QC) protocol for ED-DUB-STEP2

Samples. A total of 2,646 DNA samples for genotyping at the Broad Institute were obtained from, and subsequently analyzed as, four distinct panels, using three different Affymetrix genotyping arrays (Table S3).

Table S3: Breakdown of samples that form the ED-DUB-STEP2 study.

Panel	Cases					Controls			
	Origin	Diagnosis	Before QC	After QC	Affymetrix array	Origin	Before QC	After QC	Affymetrix array
1	Dublin	BP1	171	150	6.0	Dublin	918	799	6.0
2	Edinburgh	BP1	299	283	5.0	Edinburgh	328	275	6.0
3	STEP-UCL	BP2	639	552	5.0 (587), 500K (52)	NIMH	103	103	500K
4	STEP	BP1	113	113	500K	NIMH	90	90	500K

Samples from panel 1 were collected as part of a larger study of the genetics of psychotic disorders in the Republic of Ireland. Ethics committee approval was obtained from all participating hospitals and centers. Cases provided written informed consent and were recruited from Hospitals and Community psychiatric facilities in Ireland by a psychiatrist or psychiatric nurse trained to use the Structured Clinical Interview for DSM. Diagnosis was based on the structured interview supplemented by case note review and collateral history where available. All diagnoses were reviewed by an independent reviewer. Controls were ascertained with informed consent from the Irish GeneBank and represented blood donors who met the same ethnicity criteria as cases. Controls were not specifically screened for psychiatric illness. Individuals taking regular prescribed medication are excluded from blood donation in Ireland and donors are not financially remunerated. The Edinburgh sample, the STEP bipolar cases and the NIMH controls that form panels 2-4 were ascertained as described previously[2]. Approval for each study was obtained from the institutional reviews boards and informed consent was obtained from all subjects as described previously[2].

QC protocol. QC was performed separately for samples genotyped on the same Affymetrix platform using the PLINK toolset [3]. The QC for the 358 samples genotyped on the Affymetrix 500K array was carried out simultaneously with the STEP-UCL study as described previously[2]. We followed the same procedure for the QC of 2,303 samples genotyped on the Affymetrix 5.0 and 6.0 (Table S4). Briefly, before evaluating genotyping quality per individual, we first excluded SNPs that had a call rate < 0.90, were monomorphic or, upon genotype calling by Birdseed (http://www.broad.mit.edu/mpg/birdsuite/birdseed.html), showed gross allele frequency differences ($P < 10^{-6}$) between plates.

Table S4: Breakdown of SNP and sample filtering during QC of the ED-DUB-STEP2 samples genotyped on the Affymetrix 5.0 or 6.0 arrays.

		Affymetrix 5.0		Affymetrix 6.0	
		SNPs	Individuals	SNPs	Individuals
N at start of QC		500,568	886	906,600	1,417
N dropped at QC step:					
1	Monomorphic or with < 0.90 call rate	41,712	-	28,993	-
2	Heterozygosity or call rate < 0.95	-	35	-	107
3	Population stratification gross outliers	-	31	-	46
4	Population stratification by IBS clustering	-	31	-	0
5	Close relatives, duplications	-	0	-	38
6	Contamination	-	0	-	0
7	SNPs with call rate < 0.95	8,939	-	19,887	-
8	SNPs with HWE failure $P < 10^{-6}$	556	-	826	-
9	SNPs with MAF < 0.01	61,679	-	136,465	-
10	SNPs with PLINK chi-missing test $P < 10^{-3}$	0	-	2,657	-
11	SNPs with PLINK mishap test $P < 10^{-10}$	1,023	-	690	-
12	SNPs with a plate HWE failure $P < 10^{-6}$	492	-	501	-
N at end of QC		386,167	789	716,581	1,226

Next, we excluded individuals with heterozygosity levels > 3 standard deviations from the mean, call rate < 0.95, or very high (indicating potential kinship) or very low (indicating potential population differences) levels of average genome-wide identity-by-state (IBS) sharing with other individuals. Lastly, after excluding poorly performing samples, we removed SNPs with call rate < 0.95; Hardy-Weinberg equilibrium $P < 10^{-6}$ in cases (Affymetrix 5.0) or controls (Affymetrix 6.0); minor allele frequency < 0.01, differential call rate between cases and controls ($P < 10^{-10}$); non-random genotyping failure with respect to haplotype background ($P < 10^{-10}$), as assessed by the mishap test in PLINK; or plate-specific Hardy-Weinberg equilibrium $P < 10^{-6}$. Following these exclusions, there remained 386,167 SNPs and 789 individuals from the Affymetrix 5.0 data, and 716,581 SNPs and 1,226 individuals from the Affymetrix 6.0 data.

Following these exclusions, we merged the Affymetrix 5.0 and 6.0 data with the Affymetrix 500K data (358 samples), resulting in a total of 742,614 SNPs and 2,373 individuals. At this stage, we removed one individual from the Affymetrix 6.0 data (an Edinburgh control) with high levels of genome-wide IBS sharing with an individual from the Affymetrix 5.0 data (an Edinburgh case), and also seven individuals with high IBS sharing with samples from the STEP-UCL study.

Lastly, to account for potential chip and site effects, we further removed SNPs that had a significant ($P < 10^{-6}$) allele frequency difference between the three control groups and that were likely to represent technical artifacts rather than true population differences (n = 70); a highly significant case-control test but poor genotype clustering (n = 26); SNPs with > 10% missing data (n = 410,641), mostly those not present in all three platforms; SNPs with inconsistent strand information between arrays (n = 87); or minor allele frequency < 0.01 (n = 4).

The final ED-DUB-STEP2 dataset thus consisted of 2,365 individuals (1,098 cases and 1,267 controls) genotyped on 331,786 SNPs.

Data release policy and additional acknowledgments

Genotypes for the NIMH control samples have been submitted to the NIMH Genetics Repository and are available under the usual data release policies. Genotypes for the STEP-BD case samples will be submitted to the NIMH repository and will be available for release using the same mechanism. Researchers can apply for access to WTCCC genotype data by completing application forms available at http://www.wtccc.org.uk/.

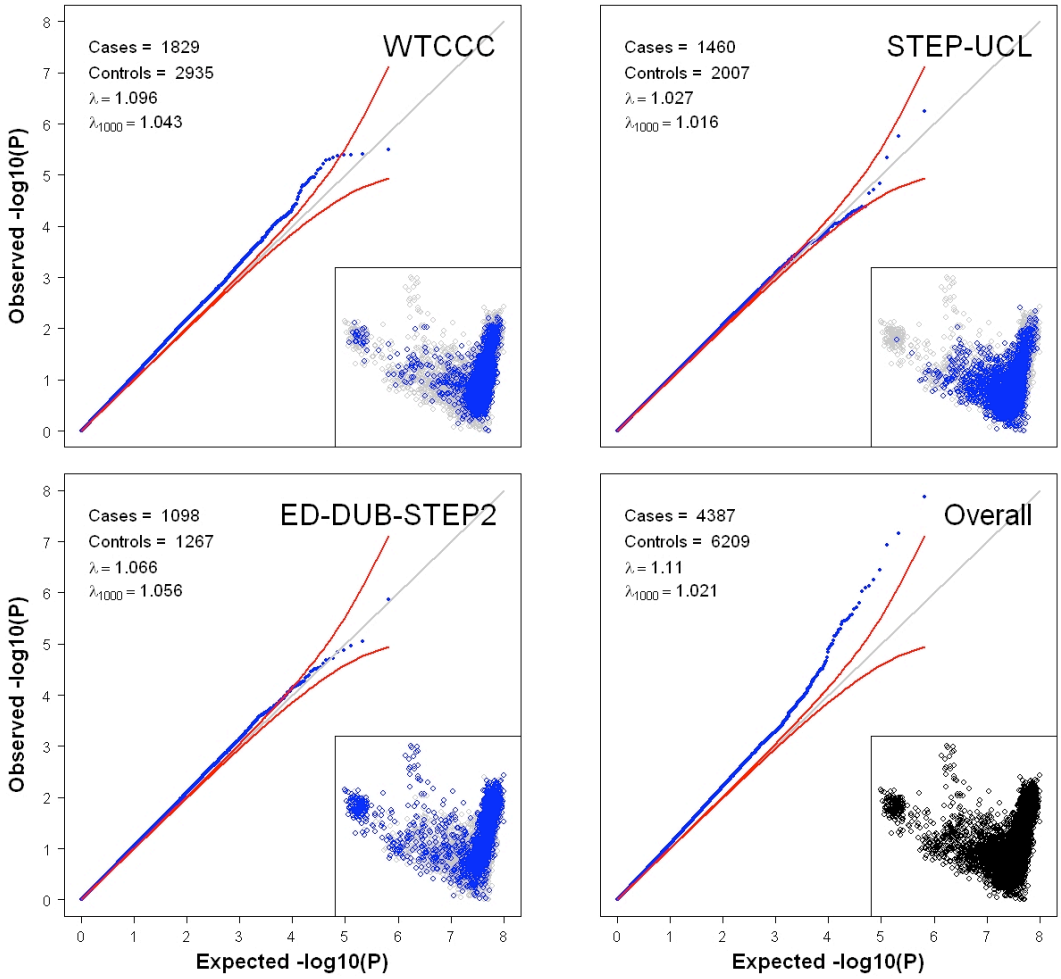
Active STEP-BD sites and principal investigators included Baylor College of Medicine (L.B. Marangell); Case University (J.R. Calabrese); Massachusetts General Hospital and Harvard

Medical School (A.A. Nierenberg); Portland VA Medical Center (P. Hauser); Stanford University School of Medicine (T.A. Ketter); University of Colorado Health Sciences Center (M. Thomas); University of Massachusetts Medical Center (J. Patel); University of Oklahoma College of Medicine (M.D. Fossey); University of Pennsylvania Medical Center (L. Gyulai); University of Pittsburgh Western Psychiatric Institute and Clinic (M.E. Thase); University of Texas Health Science Center at San Antonio (C.L. Bowden).

Control subjects from the National Institute of Mental Health Schizophrenia Genetics Initiative (NIMH-GI), data and biomaterials were collected by the ‘‘Molecular Genetics of Schizophrenia II’’ (MGS-2) collaboration. The investigators and co-investigators are as follows: ENH/Northwestern University (MH059571, P.V. Gejman, A.R. Sanders).; Emory University School of Medicine (MH59587, F. Amin); Louisiana State University Health Sciences Center (MH067257, N. Buccola); University of California-Irvine,(MH60870, W. Byerley); Washington University, St. Louis (MH060879, C.R. Cloninger); University of Iowa (MH59566, R. Crowe, D. Black); University of Colorado (MH059565, R. Freedman); University of Pennsylvania (MH061675, D. Levinson); University of Queensland (MH059588, B. Mowry); Mt. Sinai School of Medicine (MH59586, J. Silverman).

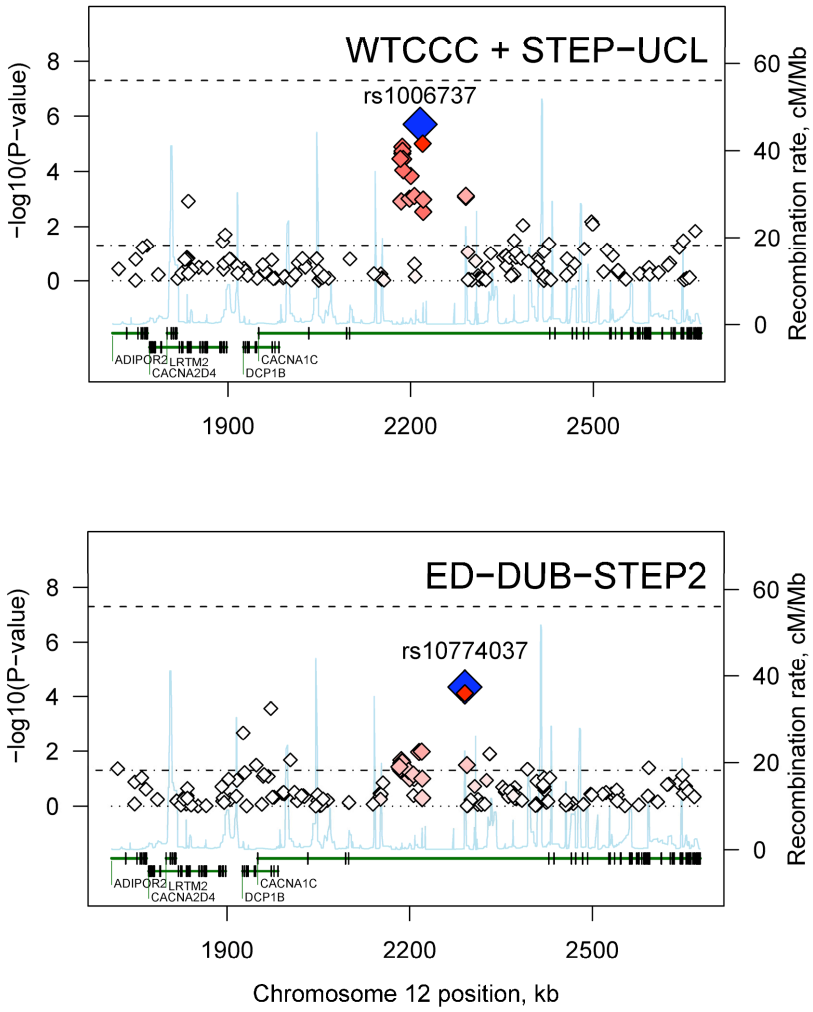
We would like to thank the individuals and families who have contributed their time and DNA to these studies. We also thank the following individuals for their assistance with this effort: F. Molay, B. Rosen-Sheidley and L. Silfies. The WTCCC study is indebted to all individuals who have participated in our research and particularly thank MDF-The Bipolar Organization for their support. The STEP-BD was supported by a Charles A. King Trust Fellowship (J.F.); NARSAD Young Investigator Awards (R.H.P., S.P., J.F.); NARSAD Independent Investigator Award (P.S.); and Johnson & Johnson Foundation (P.S.);

2. Supplementary Figures

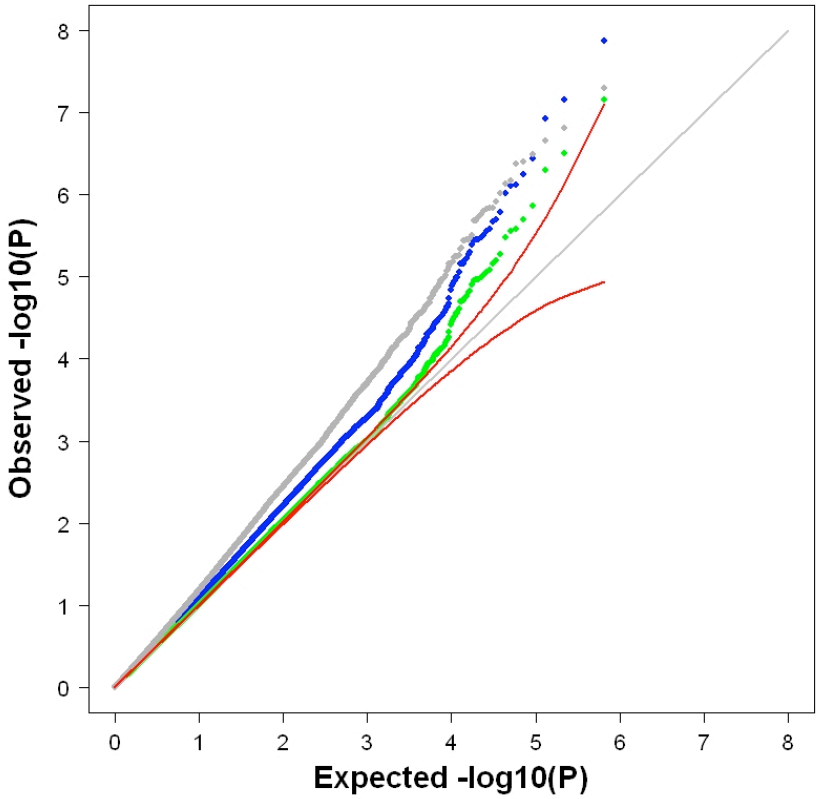


***Supplementary Figure 1*: Quantile-quantile (QQ) and multidimensional scaling (MDS) plots.** QQ plots are shown for the individual and combined datasets based on the analysis of genotyped SNPs (N = 325,690). The number of cases and controls, as well as the genomic inflation factors λ (observed) and λ_{1000} (re-scaled for a study of 1,000 cases and 1,000 controls[12-14]) are shown in the top left corner of each panel. The upper and lower boundaries of the 95% confidence bands are represented by the red lines. The insert in the bottom right corner of each graph shows a plot of the first two dimensions from a classical MDS analysis based on pairwise identity-by-state (IBS)

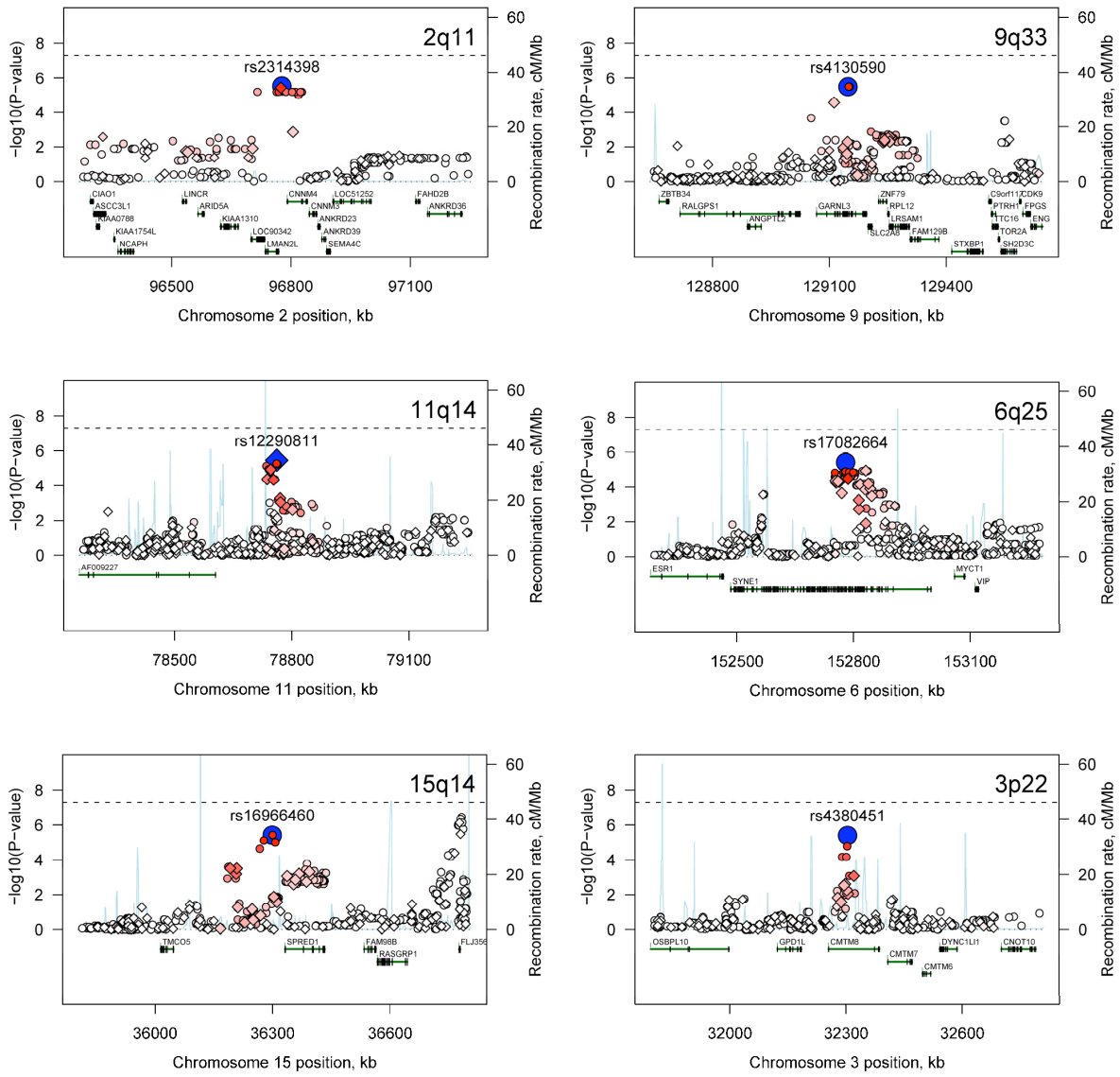
distances between pairs of individuals in the corresponding dataset. Briefly, each circle represents an individual, while individuals that cluster closely represent genetically well matched groups. For visual reference, the MDS plot for the combined dataset (insert of last graph) is shown in grey as the background for each MDS plot of the individual datasets.



***Supplementary Figure 2*: Plot of association results for the *CACNA1C* region in the combined analysis of the previously published WTCCC and STEP-UCL studies, and in the new ED-DUB-STEP2 study.** The most associated SNP is shown in blue, while the color of the remaining markers reflects the linkage disequilibrium (r^2) with the top SNP in each panel (increasing red hue associated with increasing r^2). The recombination rate (second *y*-axis) is plotted in light blue and is based on the CEU HapMap population. Horizontal lines indicate a $P = 5 \times 10^{-8}$ (dashed) and $P = 0.05$ (dash-dotted). Exons for each gene are represented by vertical bars, based on all isoforms available from the Mar 2006 UCSC genome browser assembly.

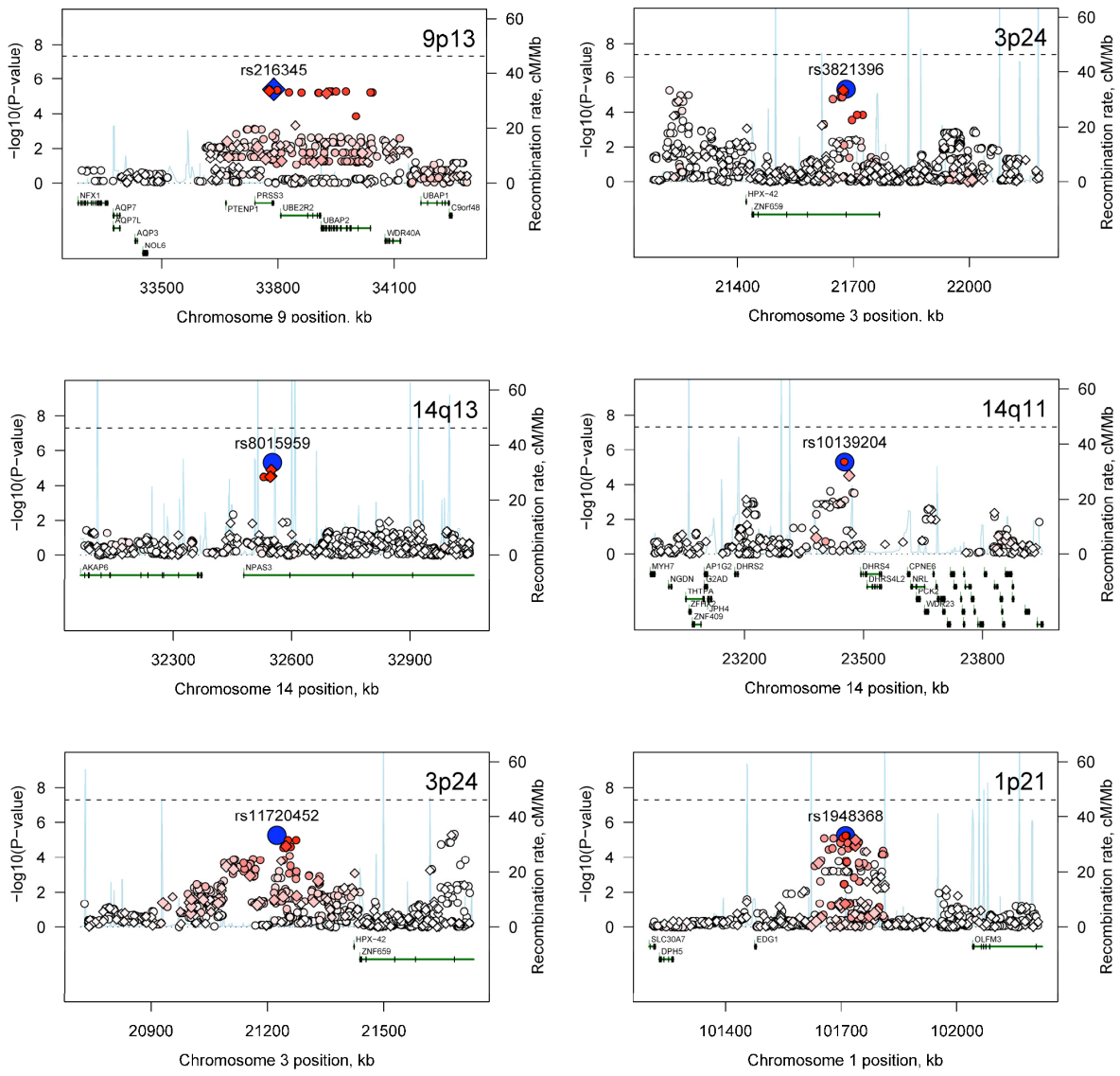


***Supplementary Figure 3*: Quantile-quantile (QQ) plot for the combined analysis.** This figure illustrates the effect of including covariates to account for study effects and two quantitative indices of ancestry (shown in blue, with $\lambda = 1.110$ and $\lambda_{1000} = 1.021$) compared to an analysis with no covariates (in grey, with $\lambda = 1.237$ and $\lambda_{1000} = 1.046$). The green curve shows the distribution of *P*-values after correction for $\lambda = 1.110$ by genomic control. Results are shown for genotyped SNPs only. The blue curve corresponds to the primary analysis reported in the main text and also displayed in the bottom-right panel of Supplementary Figure 1.



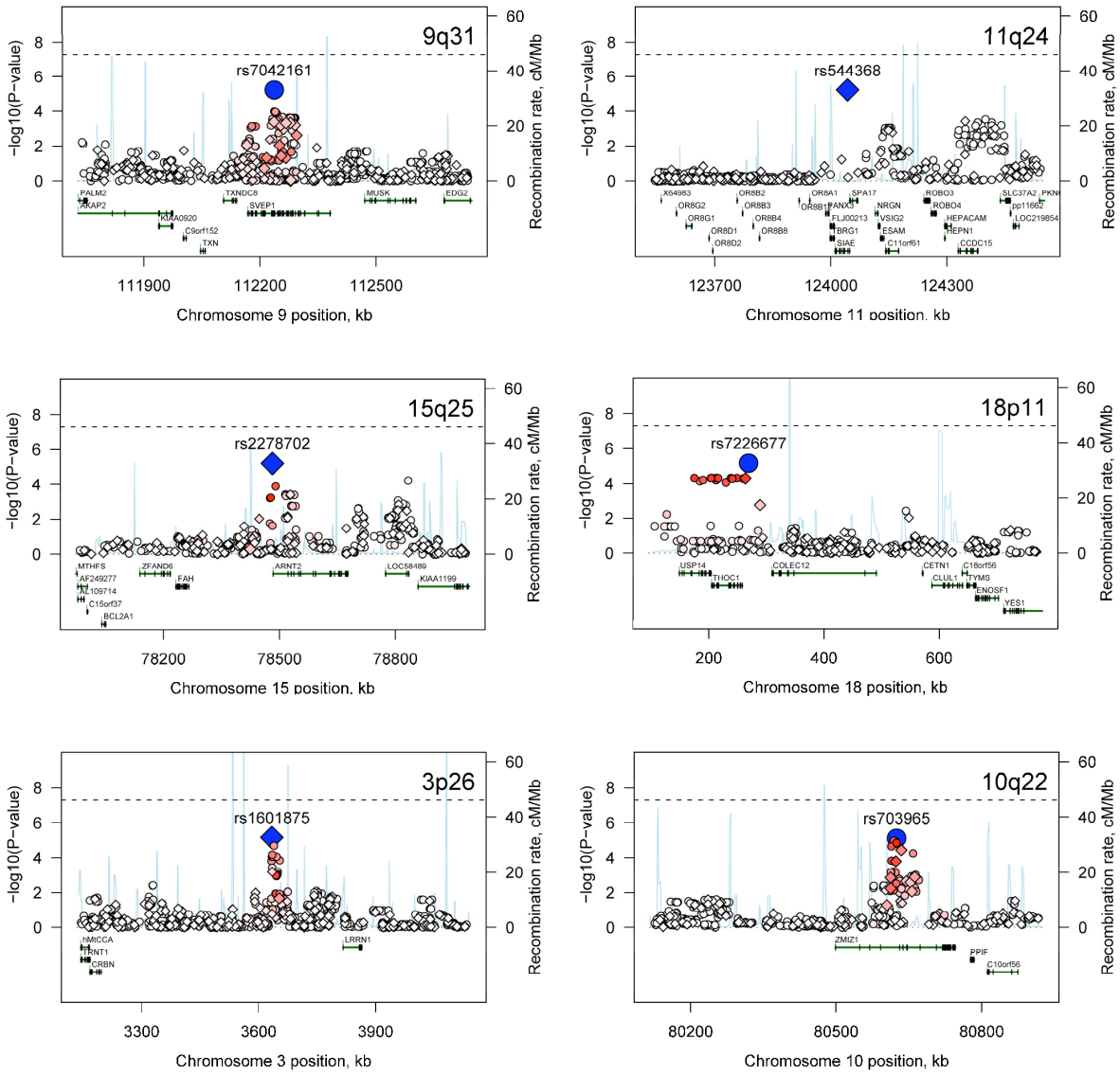
***Supplementary Figure 4*: Plots for the eighteen regions of association with $10^{-6} < P < 10^{-5}$.**

Results ($-\log_{10}P$) are shown for directly genotyped (*diamonds*) and imputed (*circles*) SNPs. The most associated SNP for each region is shown in blue, while the color of the remaining markers reflects the linkage disequilibrium (r^2) with the top SNP in each panel (increasing red hue associated with increasing r^2). The recombination rate (second *y*-axis) is plotted in light blue and is based on the CEU HapMap population. The dashed horizontal line indicates a $P = 5 \times 10^{-8}$. Exons for each gene are represented by vertical bars, based on all isoforms available from the Mar 2006 UCSC genome browser assembly.



***Supplementary Figure 4*: Plots for the eighteen regions of association with $10^{-6} < P < 10^{-5}$.**

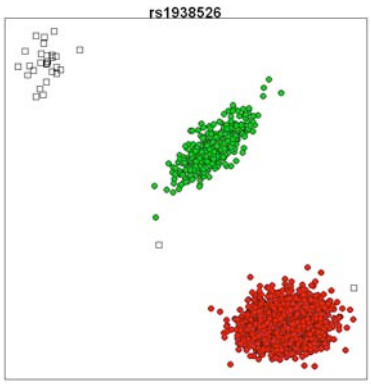
(*Continued).*

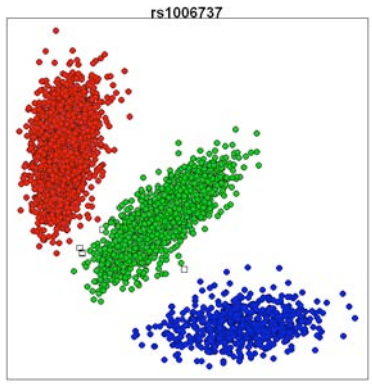


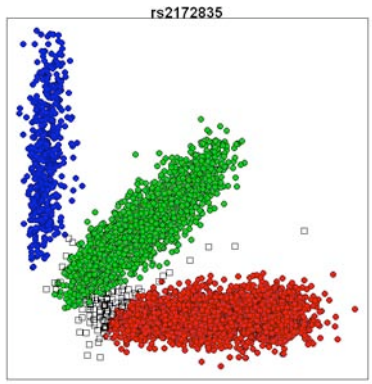
***Supplementary Figure 4*: Plots for the eighteen regions of association with $10^{-6} < P < 10^{-5}$.**

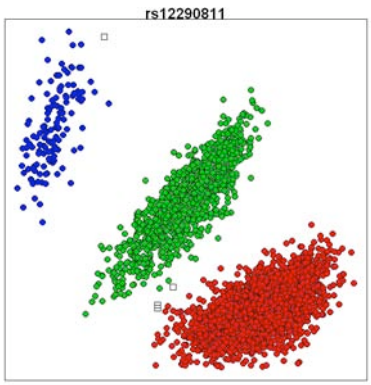
(*Continued)*

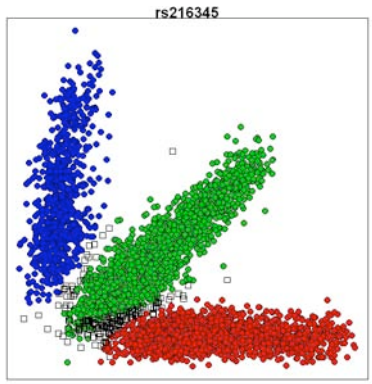
Supplementary Figure 5: Cluster plots for the genotyped SNPs listed in Table 1 of the main text and Supplementary Table 5. Data is shown separately for the WTCCC, STEP-UCL and ED-DUB-STEP2 studies. For the latter two, plates were called individually and we present plots separately for plates that used the same Affymetrix chip, namely the 500K classic (AFFY_C), 5.0 (AFFY_5.0) or 6.0 (AFFY_6.0) chips. Homozygote individuals are shown in blue (minor allele) or red (major allele), heterozygote individuals in green. Genotypes called as missing are represented by open squares. For SNP rs1938526, we manually changed the genotype calls for 27 individuals that form the homozygote cluster in the WTCCC panel and that were originally called as missing. The overall evidence for association for this SNP with and without these individuals was respectively $P = 1.3\times10^{-8}$ and 5.4×10^{-8}.

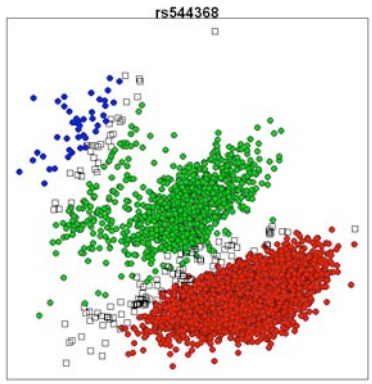

rs1938526

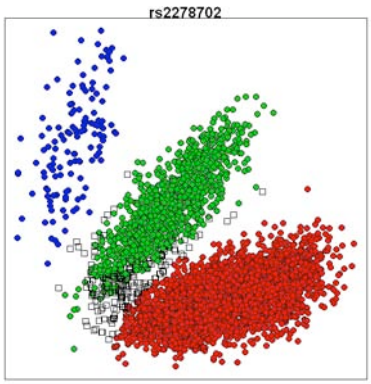

rs1006737

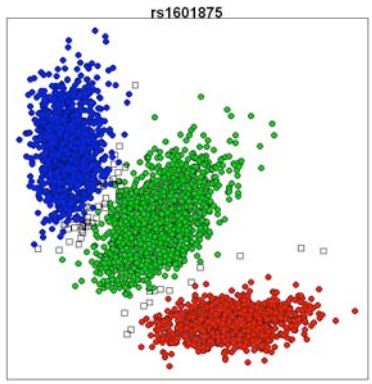

rs2172835


rs12290811


rs216345


rs544368


rs2278702

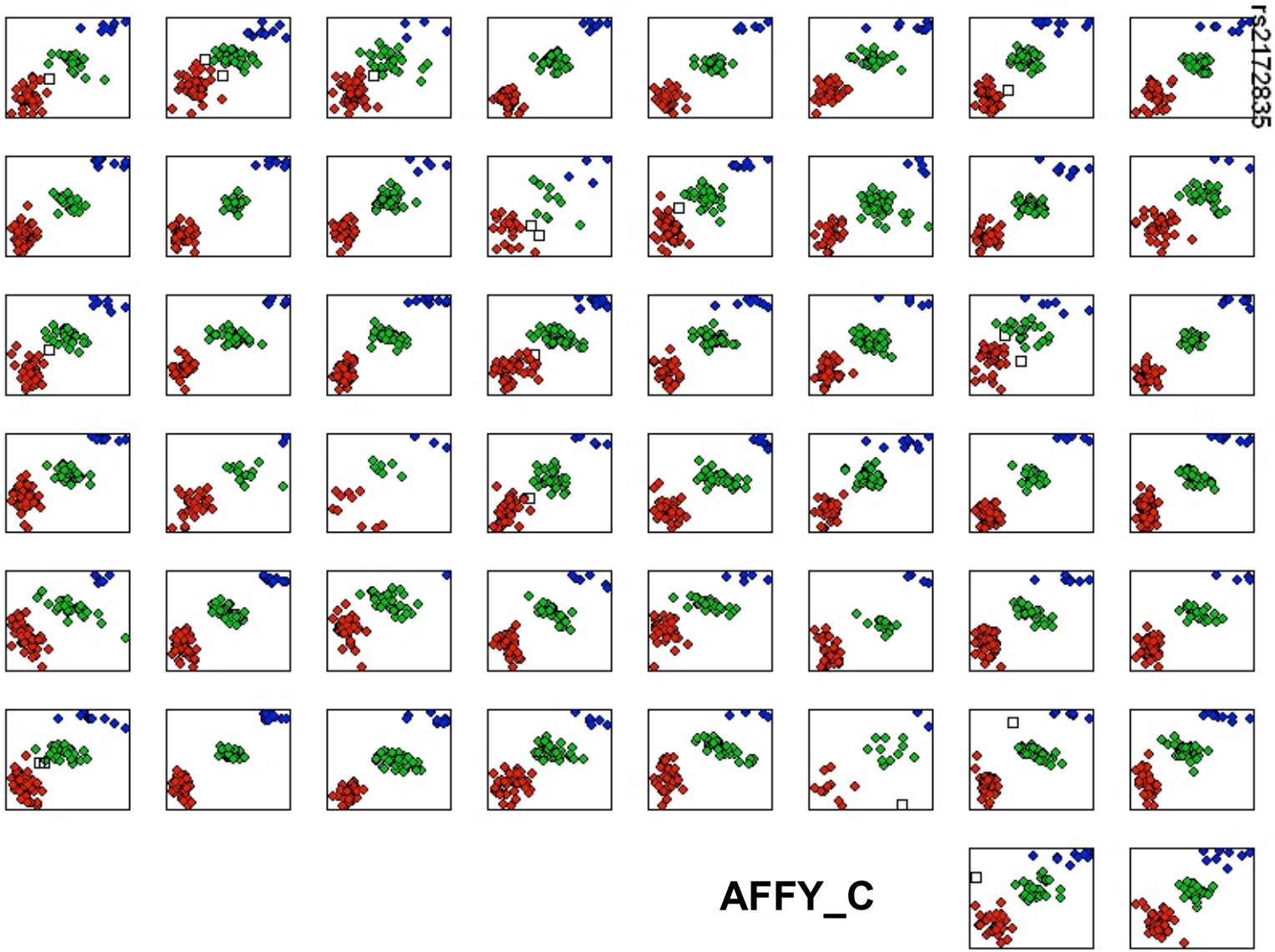

rs1601875

WTCCC

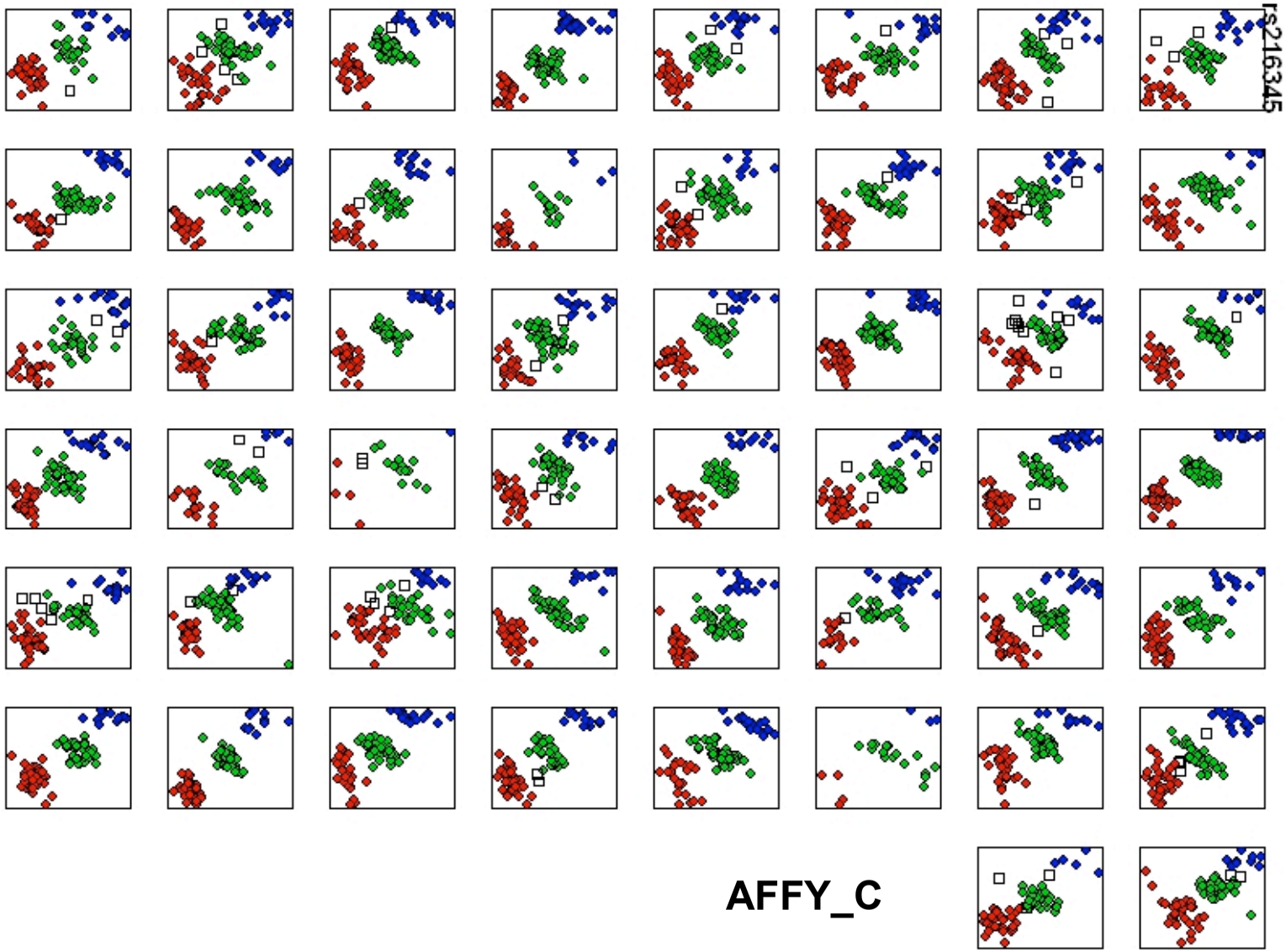
rs1938526

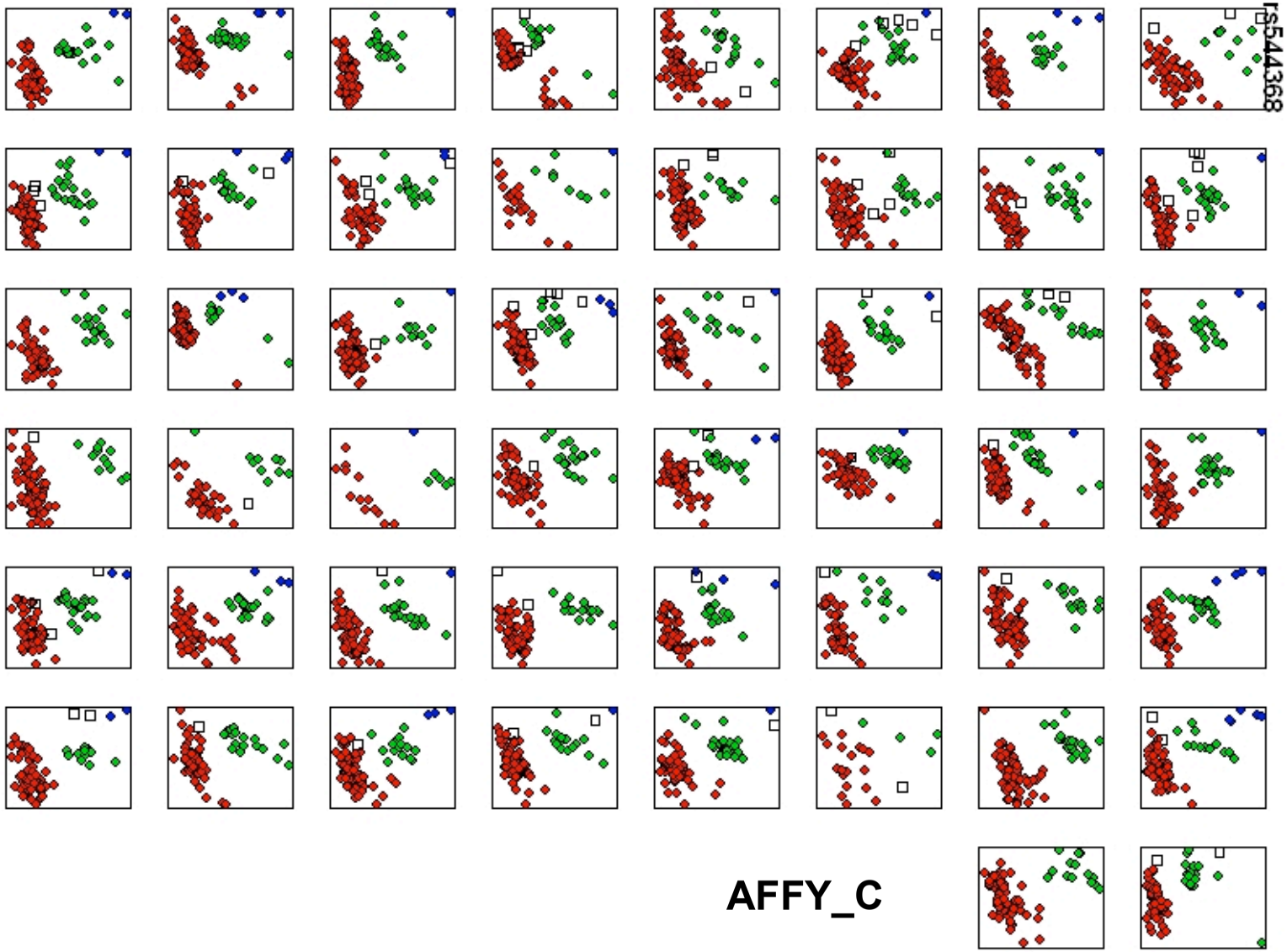
AFFY_C

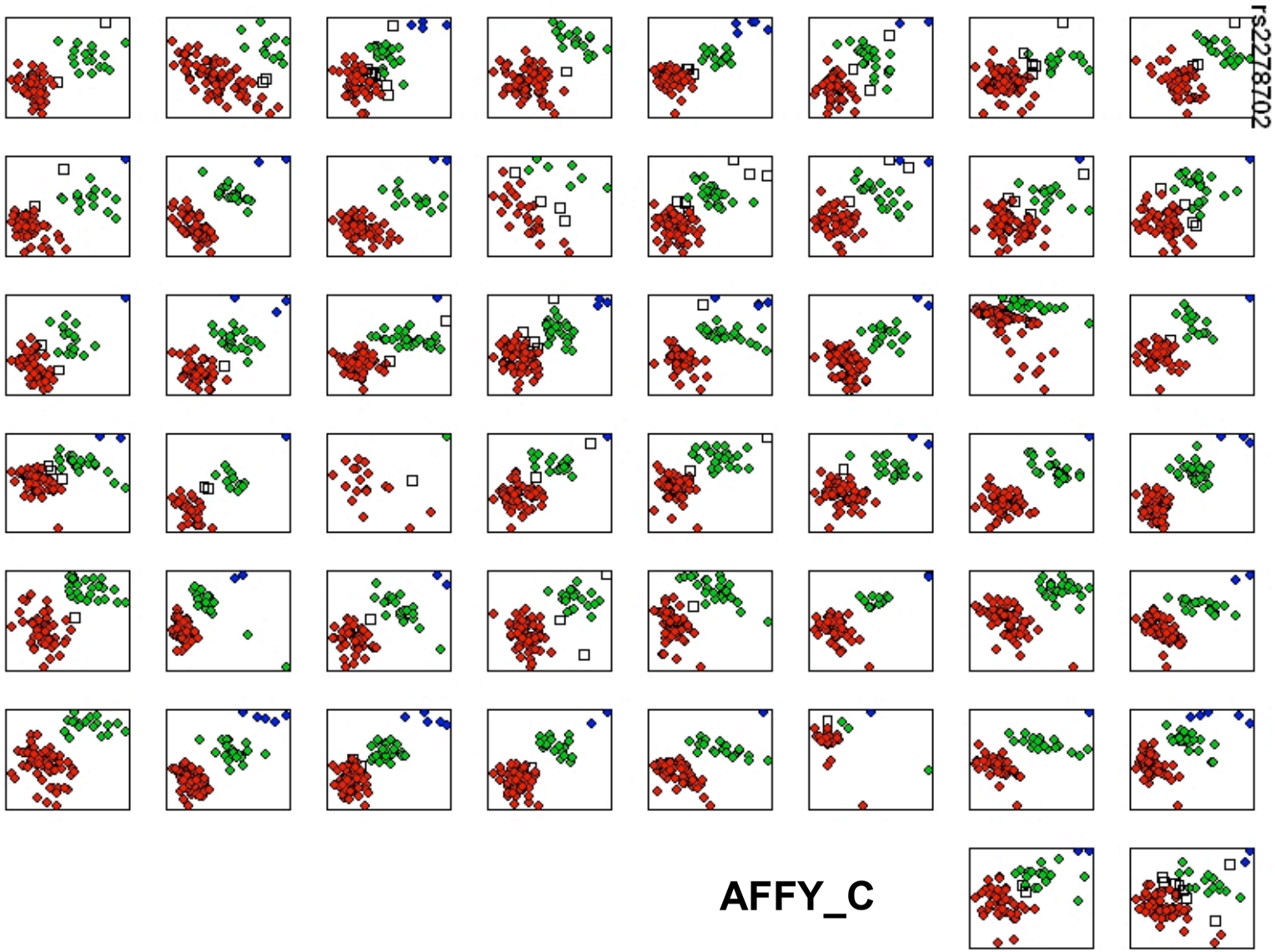
rs1006737
AFFY_C


rs2172835
AFFY_C

rs12290811
AFFY_C

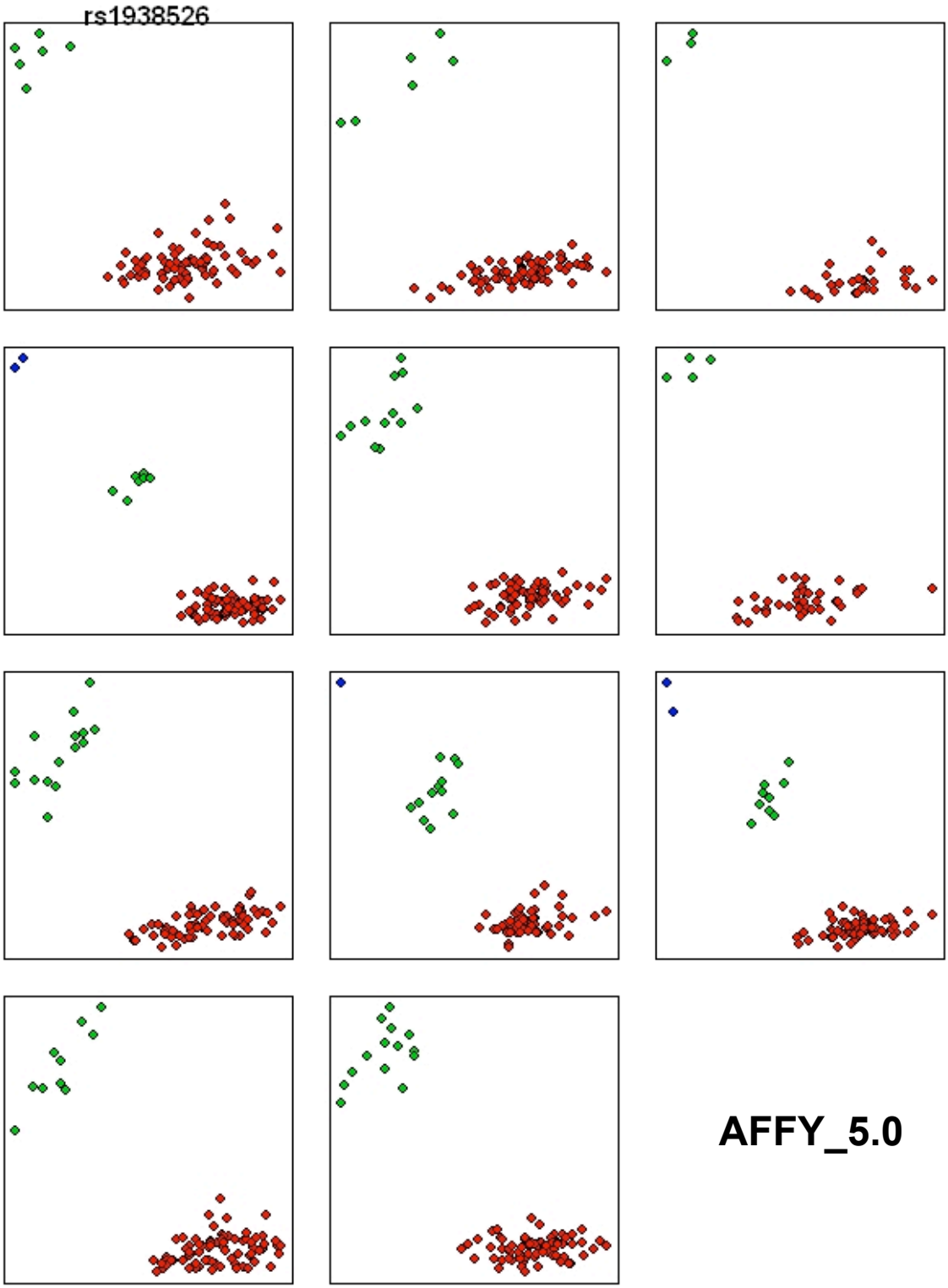

rs216345
AFFY_C

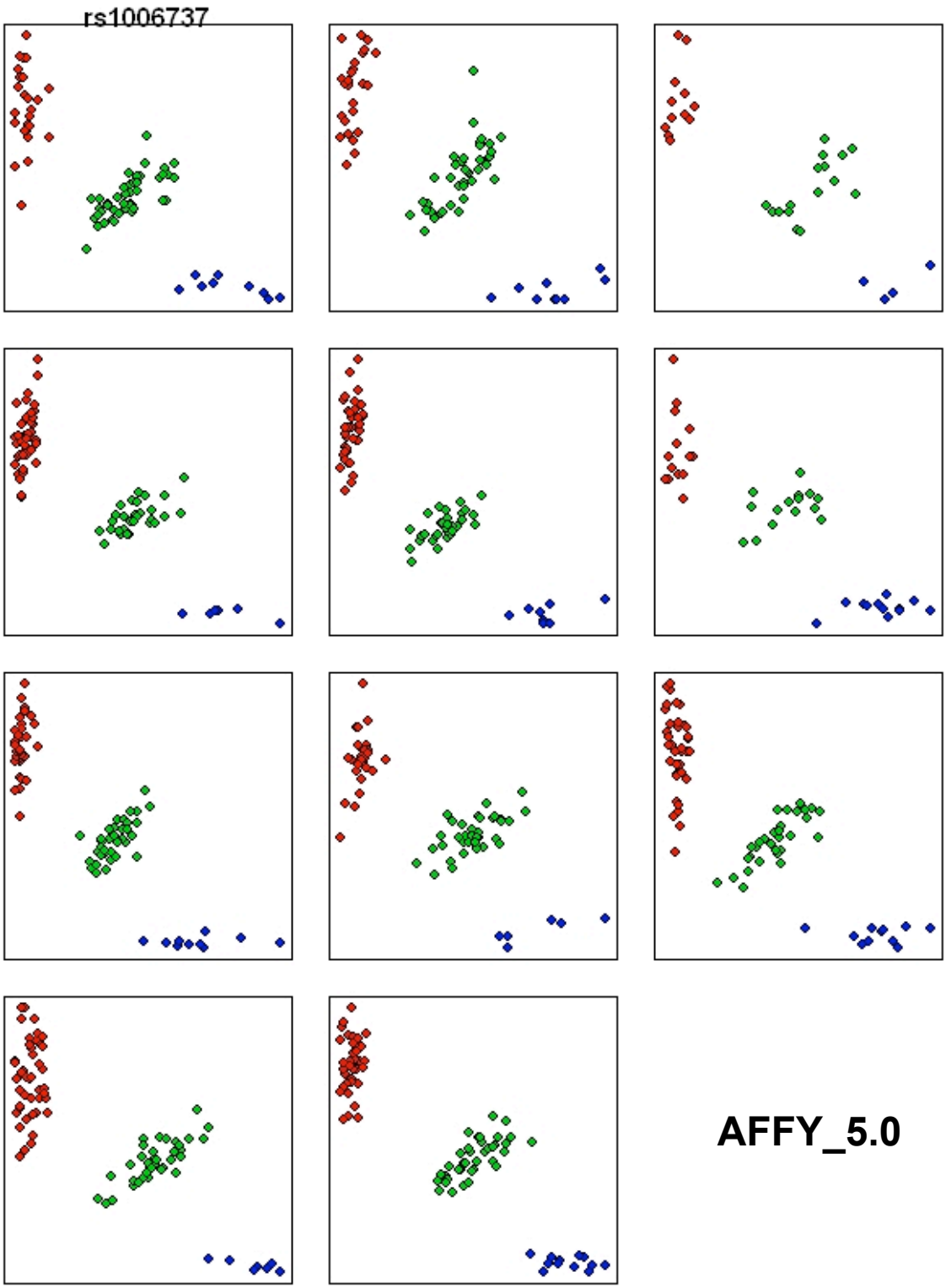

rs544368
AFFY_C

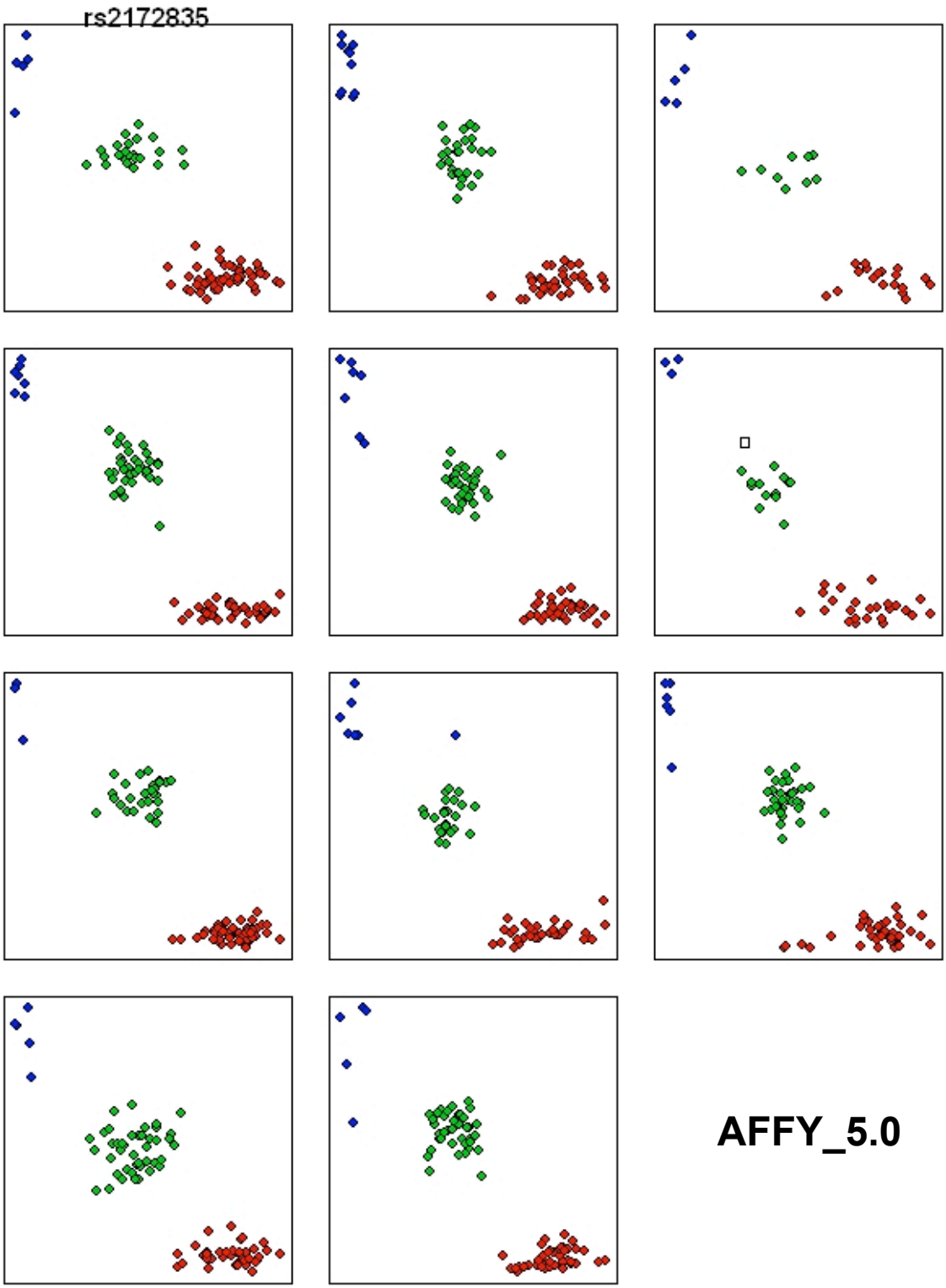

rs2278702
AFFY_C

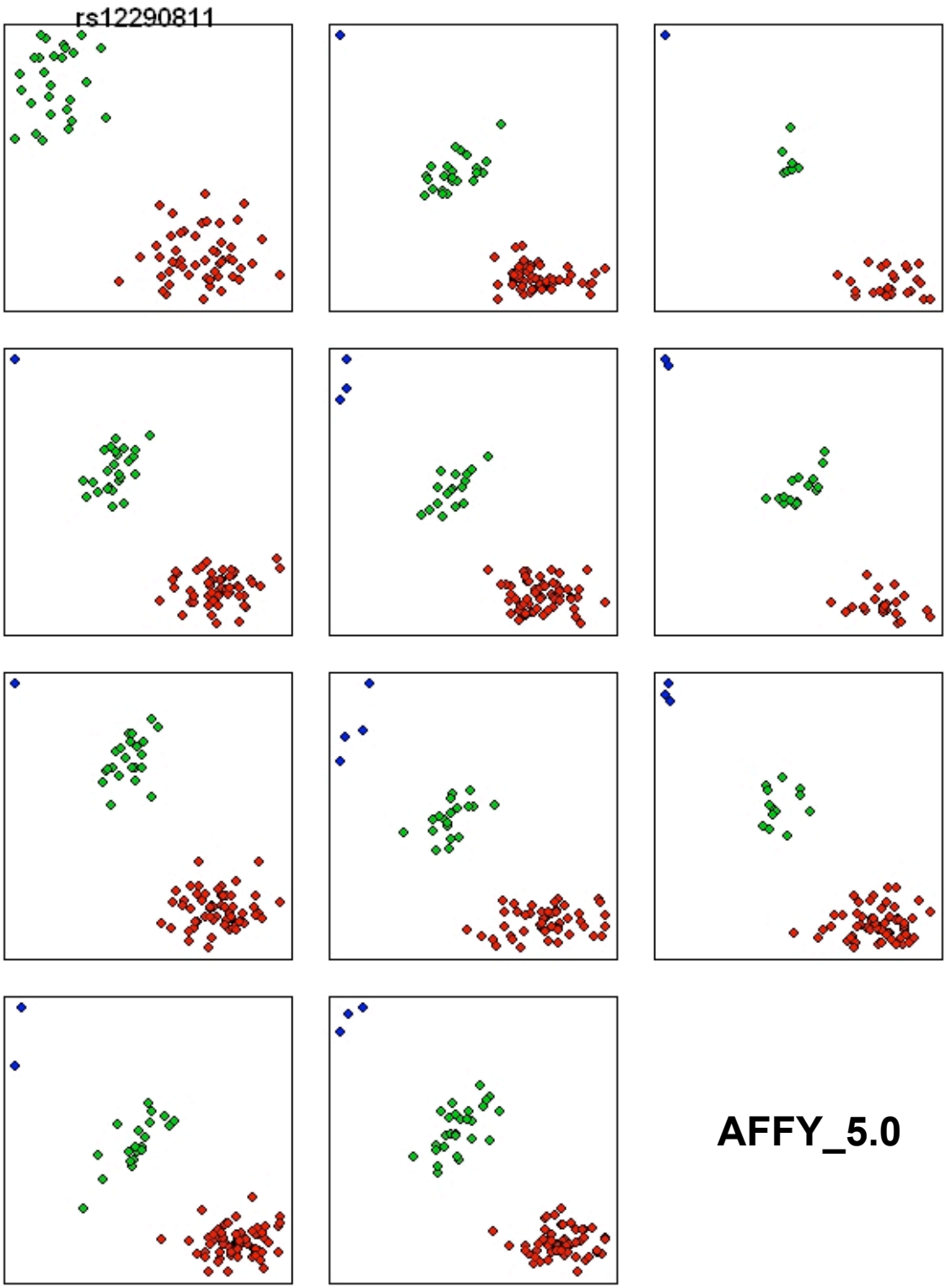
rs1601875

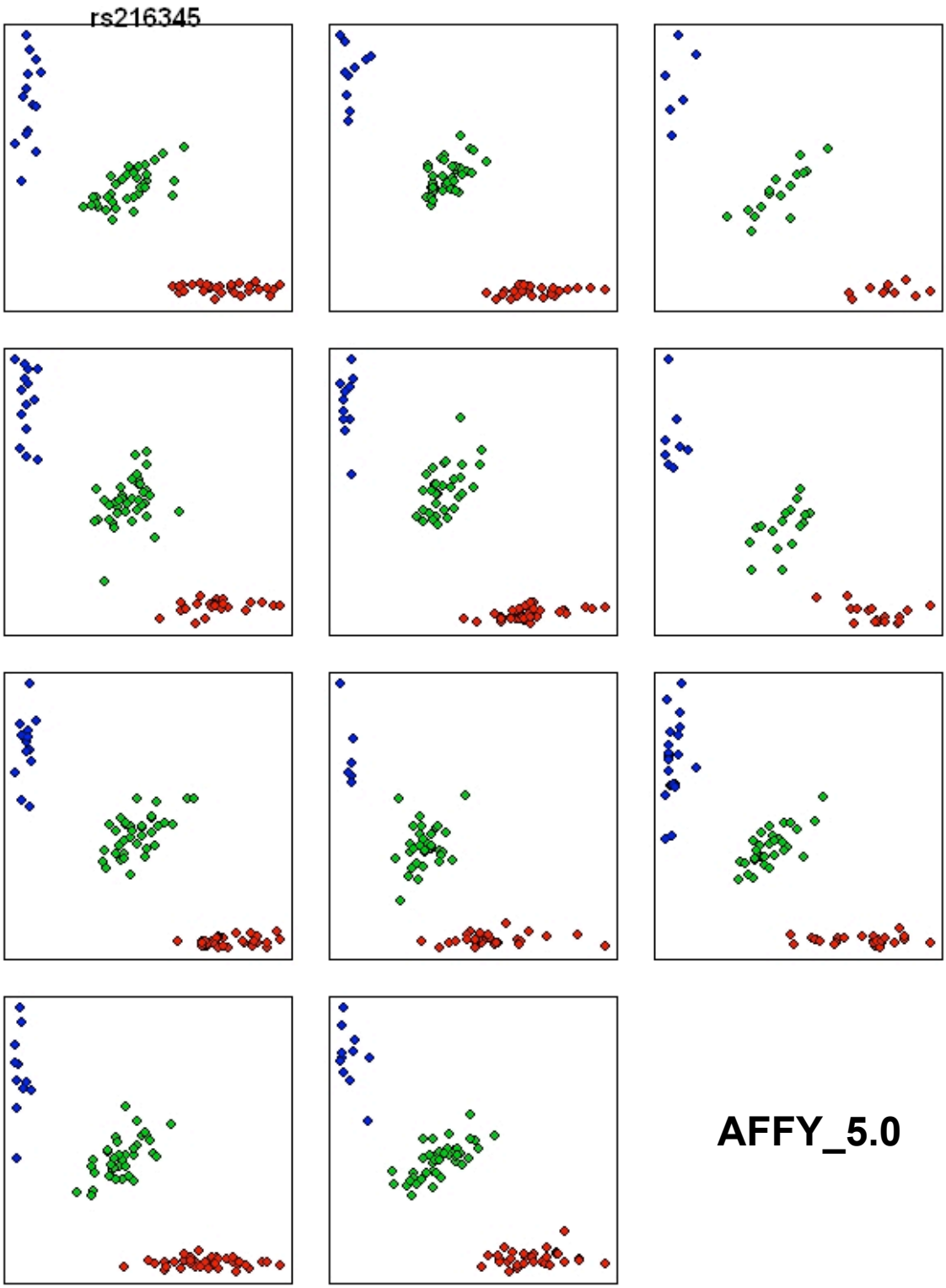
AFFY_C

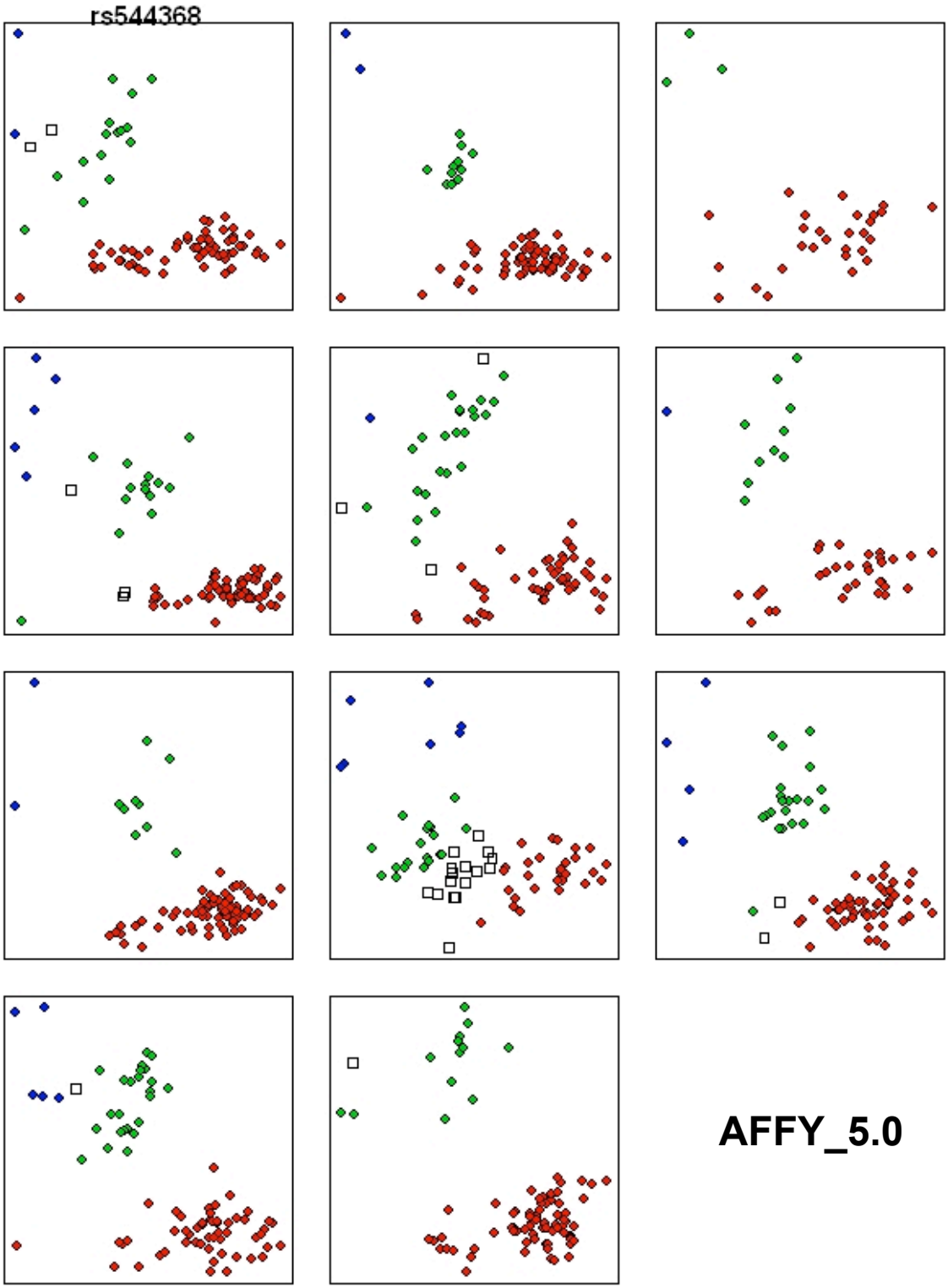

rs1938526
AFFY_5.0

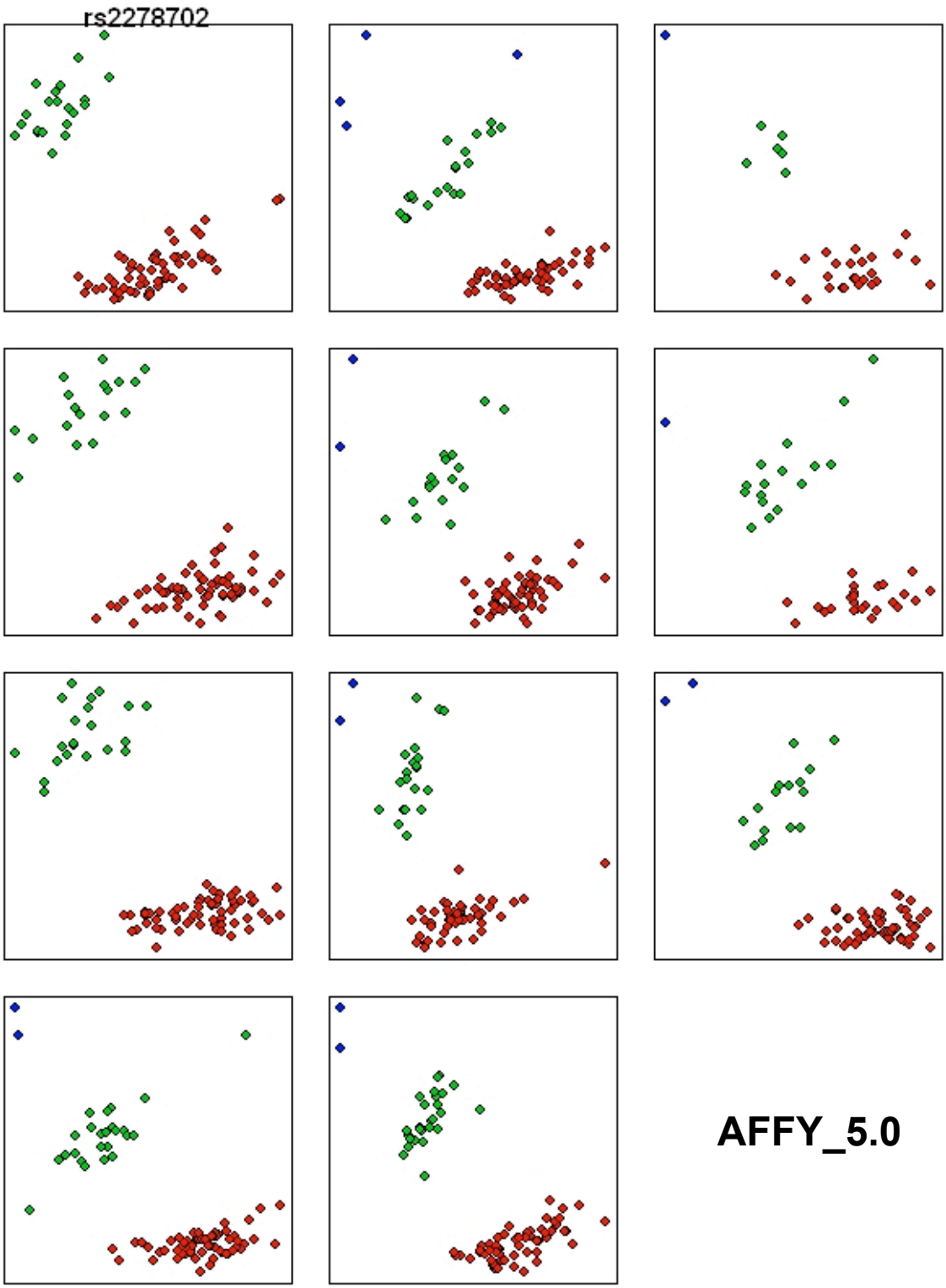

rs1006737
AFFY_5.0

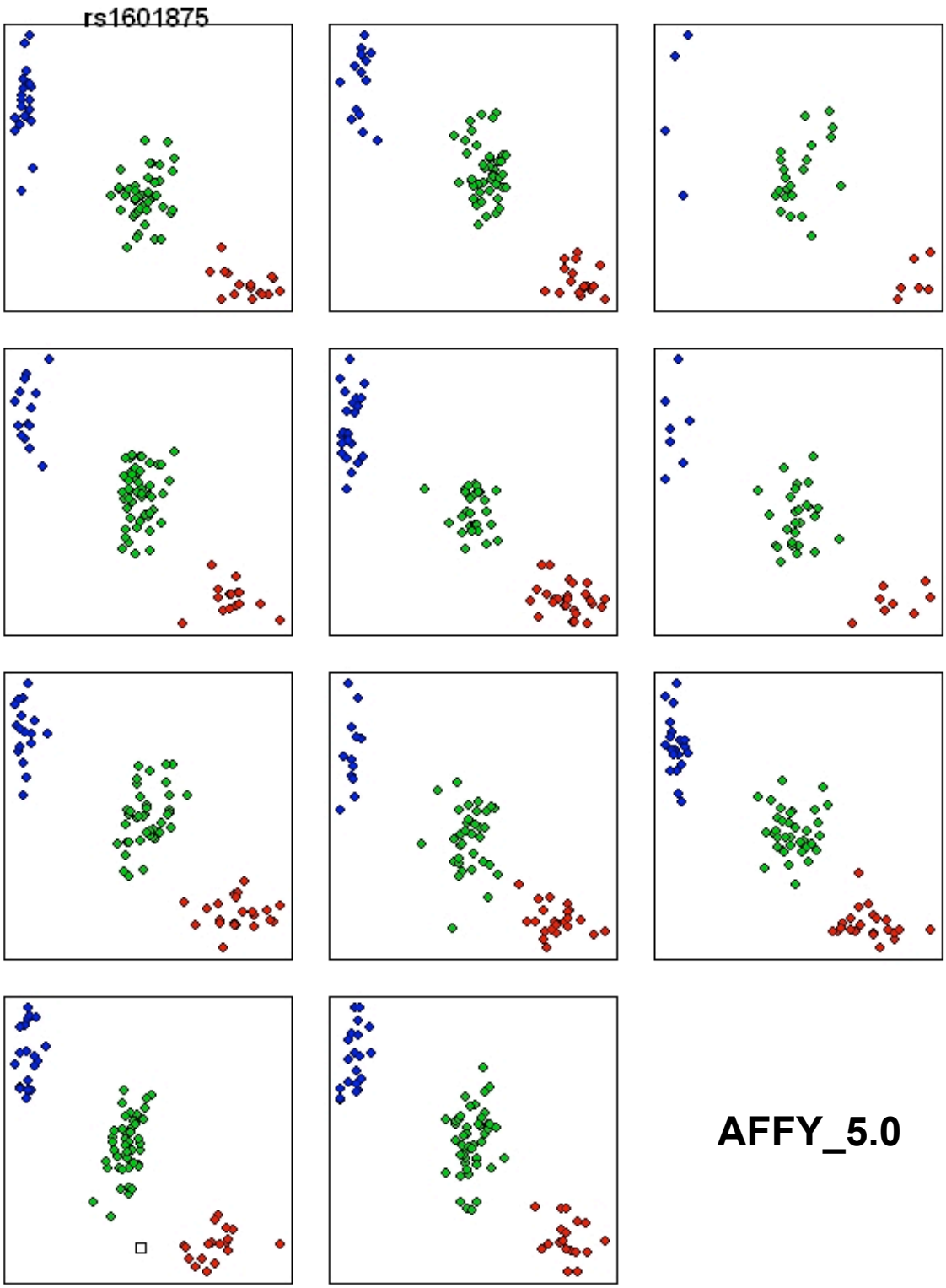

rs2172835
AFFY_5.0

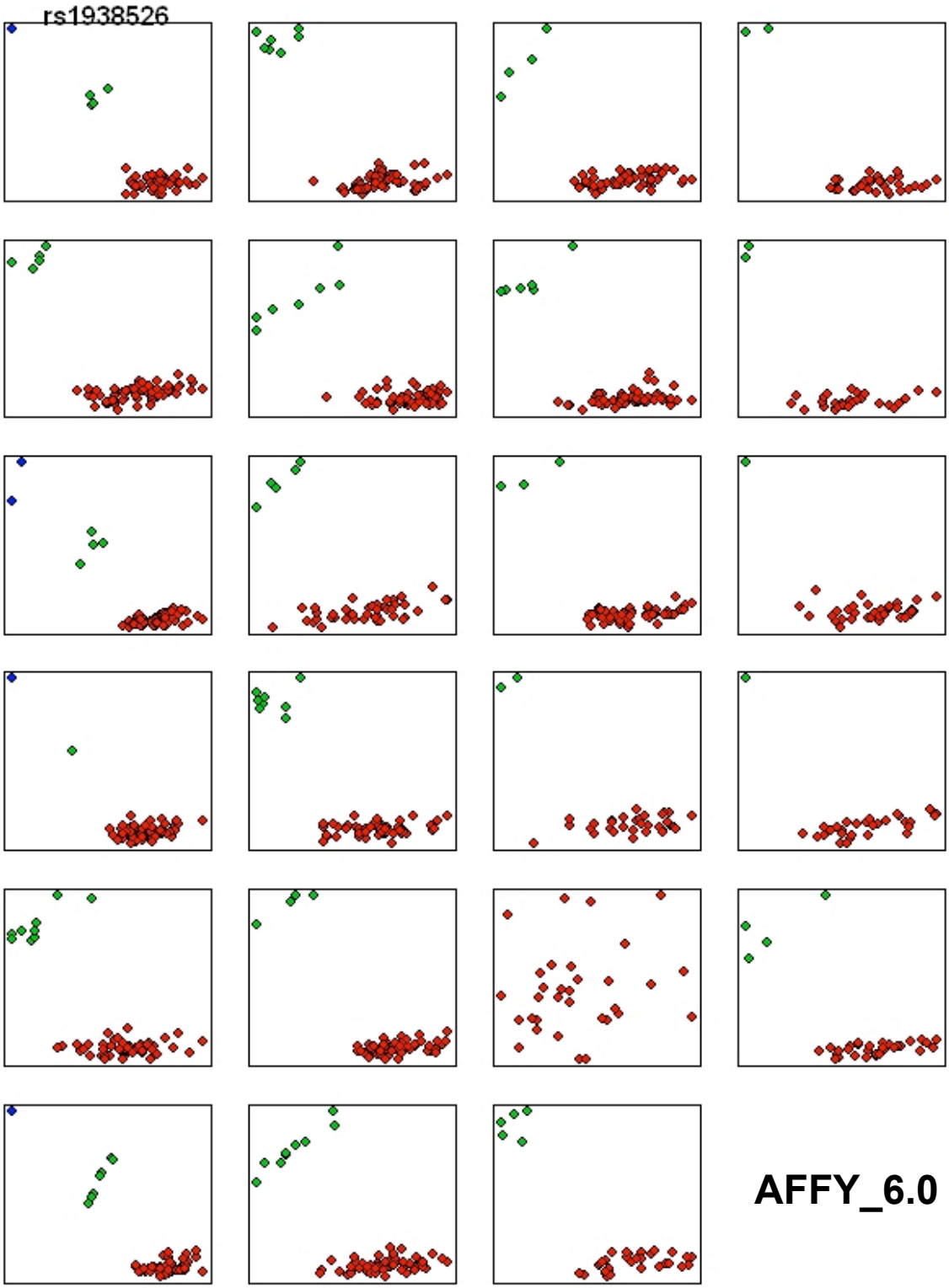

rs12290811
AFFY_5.0

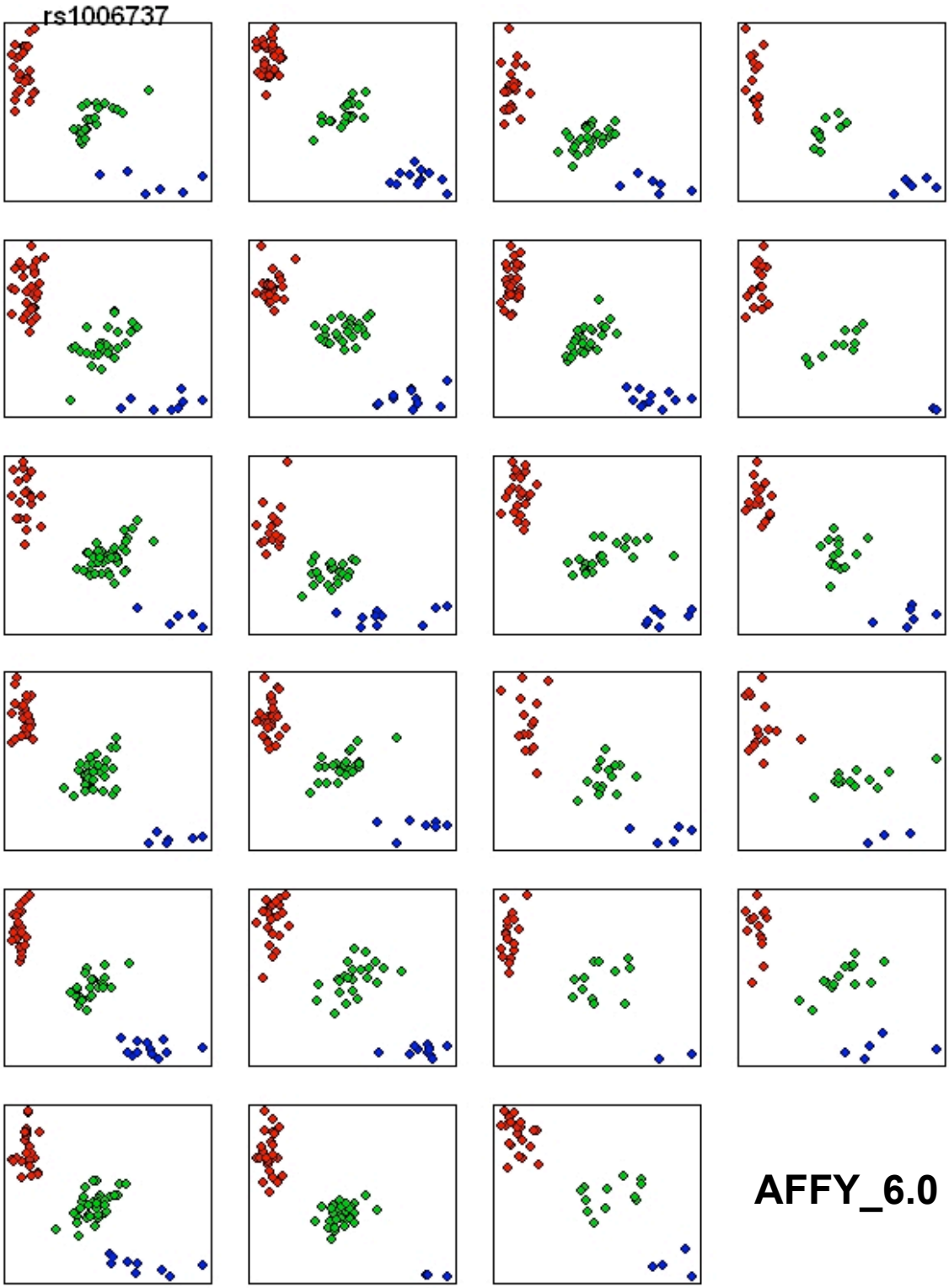

rs216345
AFFY_5.0

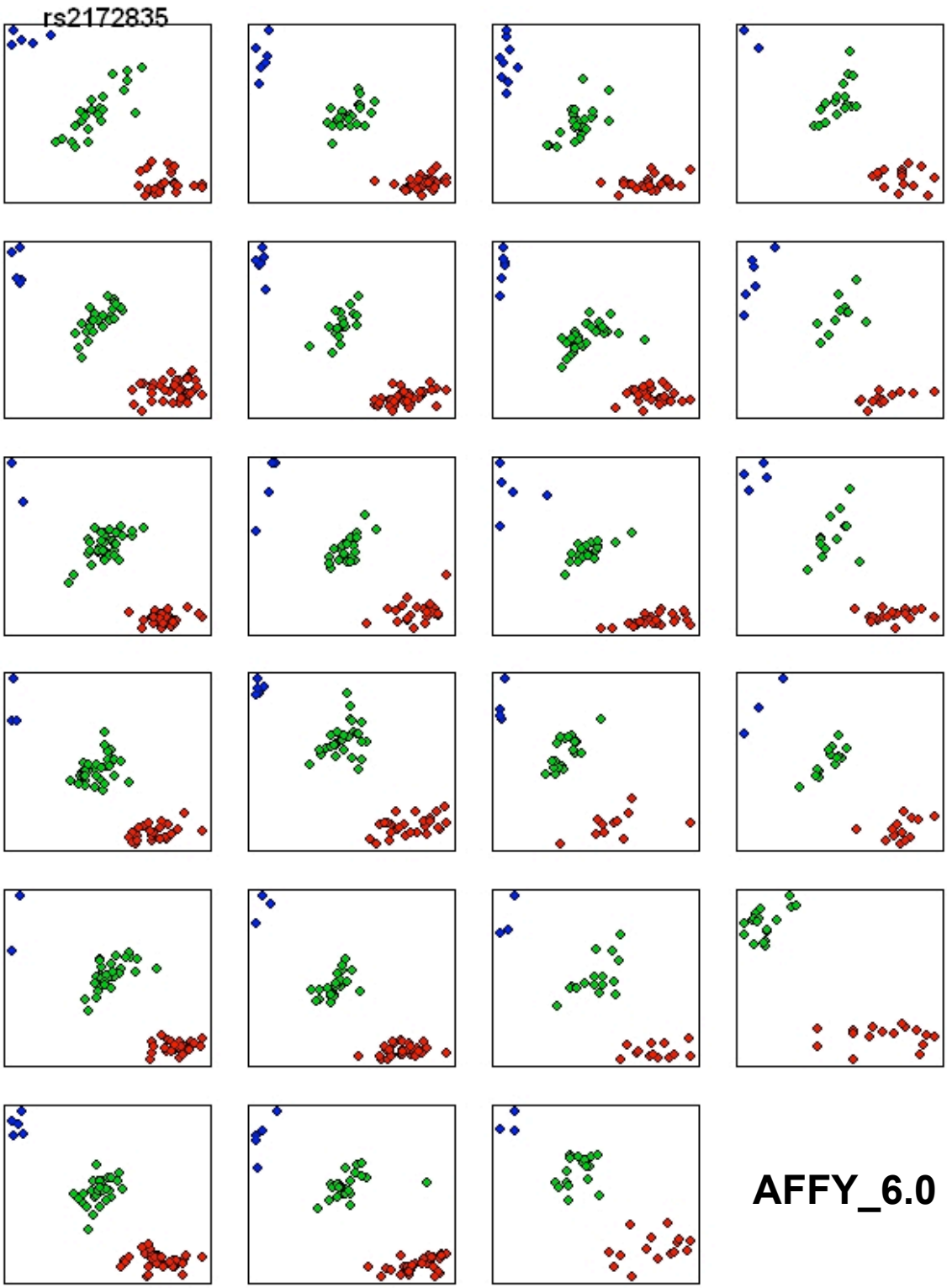

rs544368
AFFY_5.0

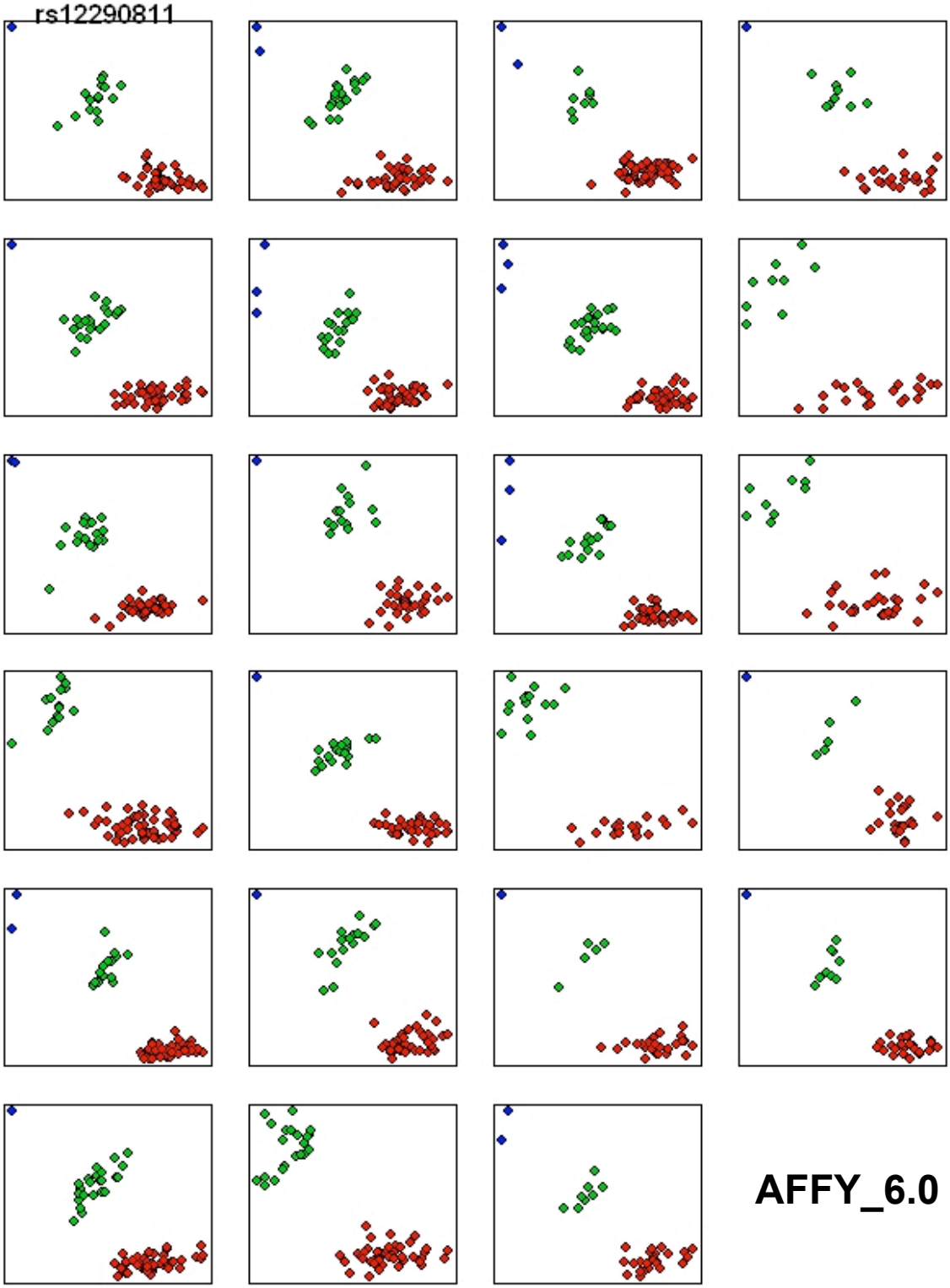

rs2278702
AFFY_5.0

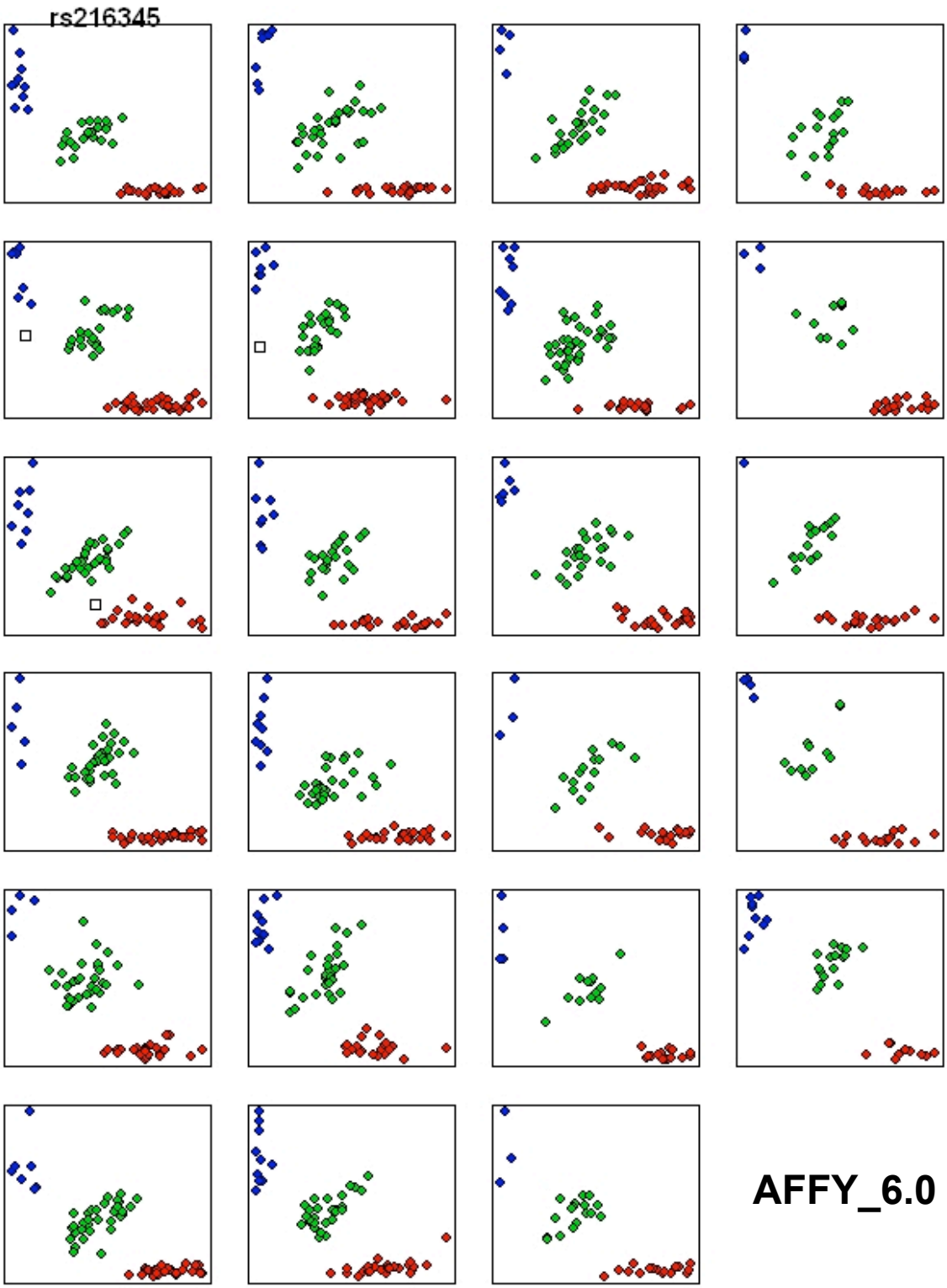

rs1601875
AFFY_5.0

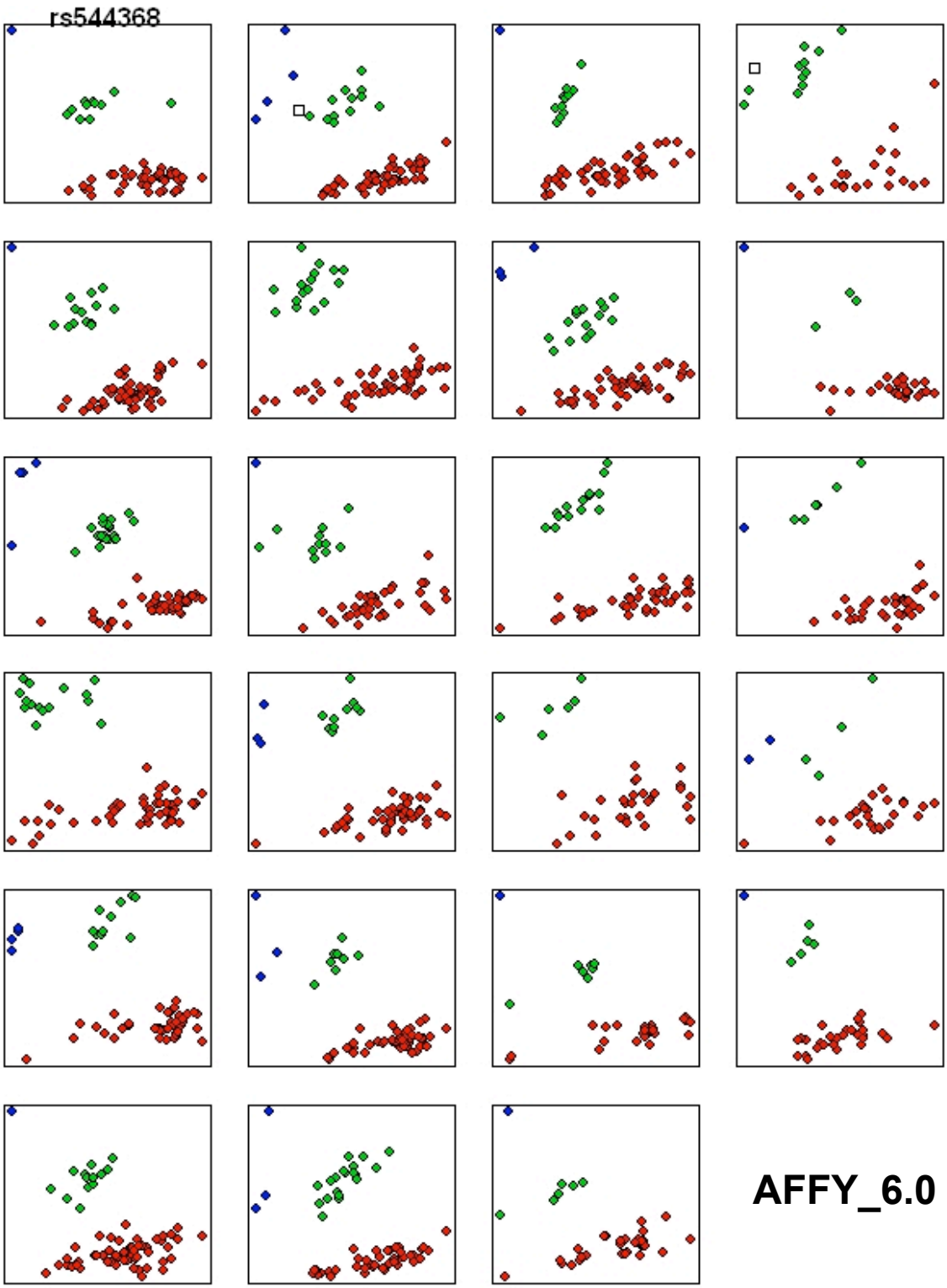

rs1938526
AFFY_6.0

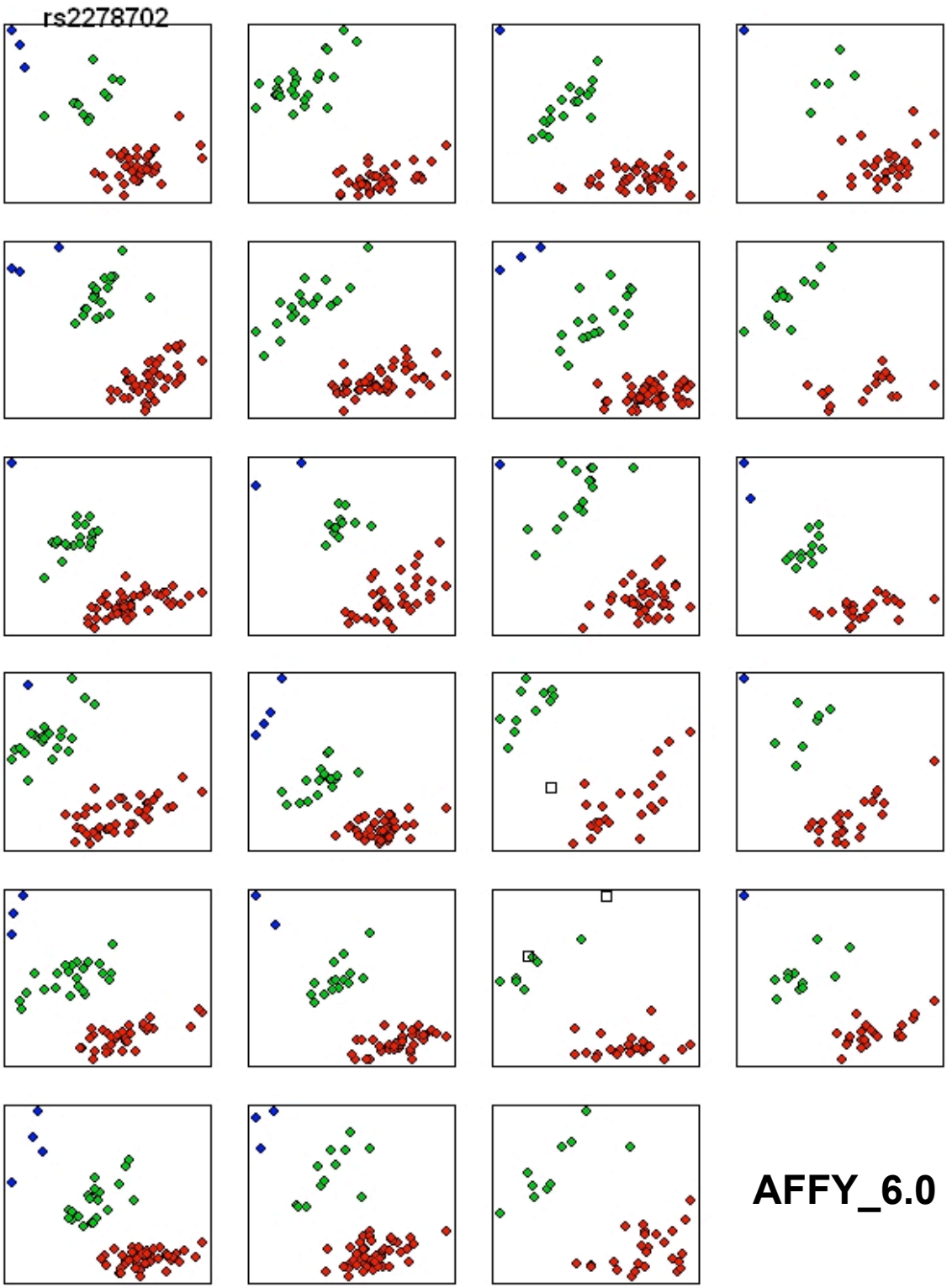

rs1006737
AFFY_6.0

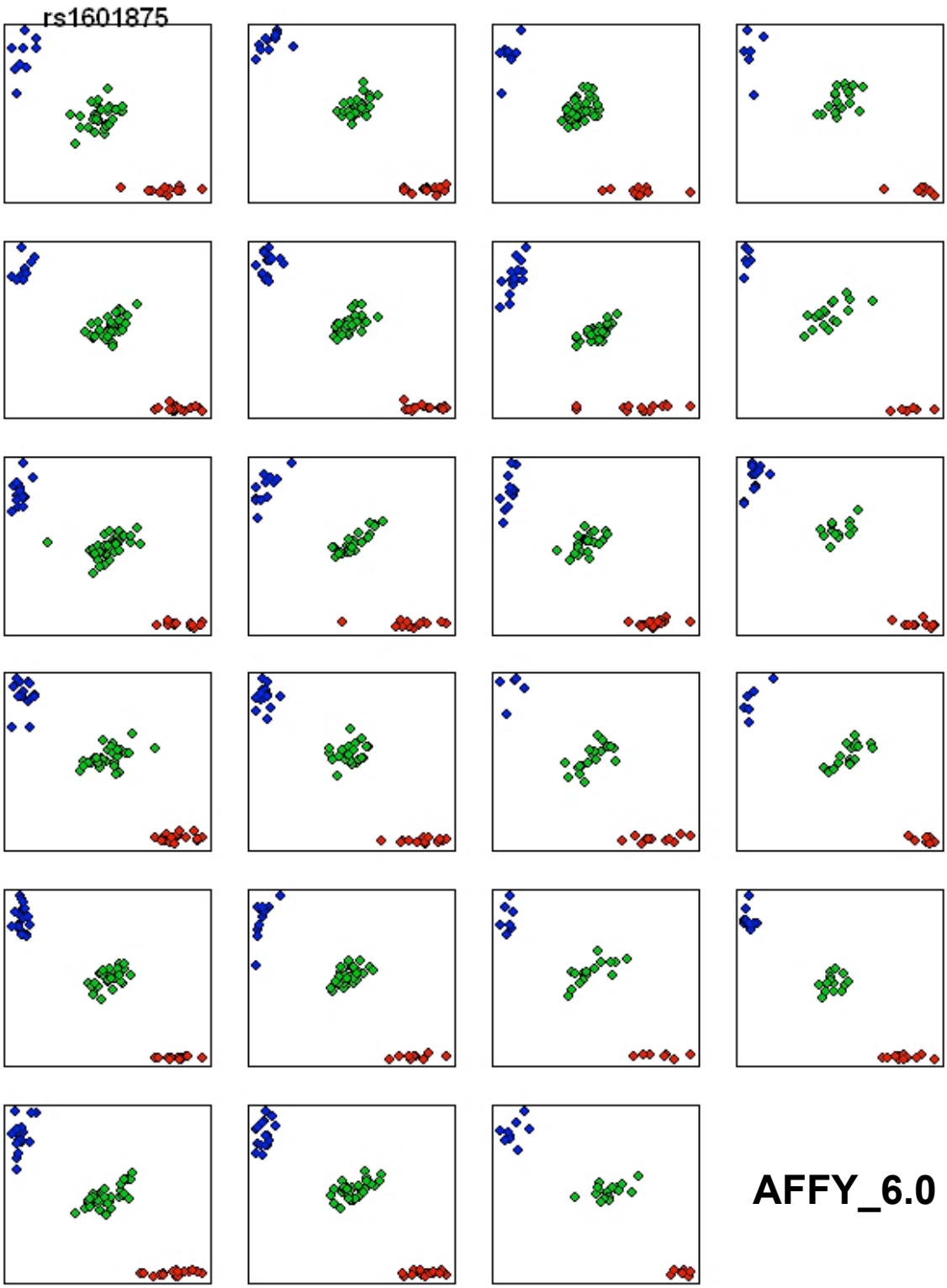

rs2172835
AFFY_6.0


rs12290811
AFFY_6.0


rs216345
AFFY_6.0


rs544368
AFFY_6.0


rs2278702
AFFY_6.0


rs1601875
AFFY_6.0

3. Supplementary Tables

Supplementary Table 1: Sample size and power for individual and combined datasets.

		WTCCC	STEP-UCL	ED-DUB-STEP2	Overall
		Sample Size			
N (% males)		4,764 (45)	3,467 (47)	2,365 (40)	10,596 (44)
Cases (% males)		1,829 (38)	1,460 (43)	1,098 (44)	4,387 (41)
Controls (% males)		2,935 (49)	2,007 (50)	1,267 (36)	6,209 (47)
Genotype missing rate		0.0027	0.0057	0.0031	0.0038
MAF	**GRR**	*Power* ($\alpha = 5 \times 10^{-8}$)			
0.05	1.40	0.05	0.02	<0.01	0.61
0.20	1.20	0.03	<0.01	<0.01	0.48
0.40	1.15	0.02	<0.01	<0.01	0.31

MAF, minor allele frequency. GRR, genotype relative risk.

Supplementary Table 2: Chromosomal regions with at least one SNP associated with bipolar disorder at $P < 5\times10^{-5}$ in the ED-DUB-STEP2 study. The evidence for association in the WTCCC and STEP-UCL samples is also provided for comparison.

Chr	Position, kb	N SNPs with:		Top SNP	*P*	Odds Ratio	Minor allele (MA)	MA frequency		WTCCC *P*	STEP-UCL *P*
		$P<5\times10^{-5}$	$P<5\times10^{-3}$					Cases	Controls		
17q21	43,729	2	12	rs7221510	1.4E-06	1.408	C	0.478	0.412	0.6363	0.6867
7p15	21,454	1	1	rs2222521	9.0E-06	2.362	C	0.047	0.026	0.6957	0.7954
13q12	25,998	2	6	rs17084299	1.1E-05	1.463	C	0.254	0.197	0.4669	0.6319
1q32	208,476	1	2	rs7550801	1.3E-05	1.691	C	0.116	0.082	0.7989	0.6366
1p35	30,134	1	7	rs10798975	1.4E-05	1.384	T	0.412	0.340	0.1067	0.0668
3p24	21,286	1	1	rs1344868	2.0E-05	2.209	G	0.051	0.027	0.3416	0.1251
13q21	64,696	2	6	rs1629705	2.0E-05	0.733	A	0.416	0.453	0.9381	0.7374
5p15	8,619	1	3	rs1157651	2.1E-05	1.468	C	0.214	0.182	0.7700	0.3667
11q14	82,983	1	22	rs635823	2.9E-05	0.727	G	0.326	0.378	0.9659	0.6843
4q31	142,432	2	7	rs6857437	2.9E-05	1.349	C	0.494	0.439	0.5325	0.6112
15q21	44,061	2	19	rs2507370	3.2E-05	0.708	T	0.239	0.284	0.2599	0.1907
16q24	84,778	1	2	rs276992	3.4E-05	1.356	G	0.397	0.317	0.1715	0.0225
22q13	47,244	1	2	rs4823800	4.1E-05	1.464	C	0.211	0.165	0.8813	0.1591
12p13	2291	1	2	rs10774037	4.6E-05	1.429	G	0.236	0.209	0.0060	0.0513

All SNPs with Hardy Weinberg test $P > 0.05$ in controls and with genotype missing rate ≤ 0.017.

Supplementary Table 3: Breakdown of bipolar disorder subtypes, psychotic symptoms and age at onset for each study. Psychosis and age-at-onset were defined as described in the Supplementary Methods.

	WTCCC (%)	**STEP-UCL (%)**	**ED-DUB-STEP2 (%)**	**Overall (%)**
Total cases	1,829	1,460	1,098	4,387
		DSM-IV diagnosis		
Bipolar 1 subtype	1,558 (85.2)	1,460 (100.0)	546 (49.7)	3,564 (81.2)
Bipolar 2 subtype	133 (7.3)	0 (0.0)	552 (50.3)	685 (15.6)
Schizoaffective manic bipolar	97 (5.3)	0 (0.0)	0 (0.0)	97 (2.2)
Bipolar NOS	41 (2.2)	0 (0.0)	0 (0.0)	41 (1.0)
		Psychotic symptoms		
Psychotic	1,045 (57.1)	1,025 (70.2)	373 (34.0)	2,443 (55.7)
Non-psychotic	385 (21.1)	413 (28.3)	574 (52.3)	1,372 (31.3)
Unavailable	399 (21.8)	22 (1.5)	151 (13.7)	572 (13.0)
		Age at onset		
Over 18	911 (49.8)	647 (44.3)	428 (39.0)	1,986 (45.3)
18 or under	342 (18.7)	719 (49.3)	473 (43.1)	1,534 (35.0)
Unavailable	576 (31.5)	94 (6.4)	197 (17.9)	867 (19.7)

Supplementary Table 4: PLINK imputation performance based on the comparison between imputed and observed genotypes for 311,233 genotyped autosomal SNPs. The overall average concordance for SNPs imputed with an information score ≥0.8 was 0.987.

MAF	N SNPs	Imputation Info score	Proportion of SNPs	Average Imputation Rate	Average Concordance
0.01-0.05	27,078	Not imputed	0.000	-	-
		0-0.5	0.325	0.841	0.966
		0.5-0.8	0.149	0.917	0.979
		≥0.8	0.526	0.992	0.992
0.05-0.15	71,984	Not imputed	0.002	-	-
		0-0.5	0.164	0.525	0.934
		0.5-0.8	0.175	0.750	0.961
		≥0.8	0.659	0.967	0.989
0.15-0.25	65,918	Not imputed	0.004	-	-
		0-0.5	0.082	0.248	0.874
		0.5-0.8	0.164	0.554	0.939
		≥0.8	0.750	0.939	0.986
0.25-0.50	146,253	Not imputed	0.004	-	-
		0-0.5	0.053	0.094	0.777
		0.5-0.8	0.145	0.389	0.907
		≥0.8	0.798	0.917	0.981

MAF, minor allele frequency.

Supplementary Table 5: Chromosomal regions with at least one SNP associated with bipolar disorder at $P < 10^{-5}$ in the combined analysis of WTCCC, STEP-UCL and ED-DUB-STEP2 datasets.

Chr	SNP, minor allele (MA)	SNP type	Combined analysis				WTCCC			STEP-UCL			ED-DUB-STEP2		
			P	OR	MA frequency		*P*	MA frequency		*P*	MA frequency		*P*	MA frequency	
					Case	Control		Case	Control		Case	Control		Case	Control
10q21	rs10994336,T	Imputed	9.1E-09	1.450	0.070	0.053	0.0006	0.070	0.054	0.0004	0.073	0.056	0.0002	0.070	0.049
12p13	rs1006737,A	Genotyped	7.0E-08	1.181	0.356	0.324	0.0015	0.357	0.324	0.0003	0.357	0.315	0.0108	0.353	0.337
15q14	rs12899449,G	Imputed	3.5E-07	0.836	0.246	0.276	0.0140	0.253	0.277	0.0005	0.237	0.276	0.0013	0.247	0.273
2q11	rs2314398,G	Imputed	2.8E-06	0.854	0.277	0.309	0.0075	0.286	0.316	0.0388	0.278	0.306	8.E-05	0.264	0.304
9q33	rs4130590,A	Imputed	3.1E-06	0.862	0.414	0.445	4.E-05	0.413	0.453	0.1220	0.413	0.427	0.0456	0.416	0.453
11q14	rs12290811,A	Genotyped	3.6E-06	1.203	0.171	0.149	0.0025	0.173	0.148	0.0005	0.177	0.144	0.2945	0.160	0.160
6q25	rs17082664,G	Imputed	3.6E-06	1.220	0.145	0.126	0.0058	0.151	0.131	0.0003	0.157	0.129	0.3580	0.137	0.129
15q14	rs16966460,G	Imputed	3.7E-06	1.260	0.121	0.103	0.0043	0.119	0.103	0.0233	0.125	0.108	0.0018	0.119	0.094
3p22	rs4380451,T	Imputed	3.9E-06	0.845	0.243	0.268	5.E-06	0.241	0.278	0.1920	0.247	0.259	0.1130	0.241	0.261
9p13	rs216345,T	Genotyped	4.1E-06	1.150	0.406	0.372	0.0014	0.405	0.375	0.0063	0.410	0.378	0.0671	0.403	0.355
3p24	rs3821396,A	Imputed	4.6E-06	1.230	0.124	0.108	1.E-05	0.131	0.104	0.0048	0.133	0.113	0.8780	0.102	0.108
14q13	rs8015959,T	Imputed	4.7E-06	1.590	0.030	0.021	0.0163	0.031	0.023	0.0005	0.036	0.023	0.0952	0.036	0.030
14q11	rs12436436,C	Imputed	4.9E-06	1.300	0.099	0.081	0.0003	0.103	0.084	0.0057	0.099	0.083	0.2230	0.092	0.072
3p24	rs11720452,T	Imputed	5.3E-06	0.869	0.389	0.422	0.0364	0.396	0.417	0.0004	0.380	0.419	0.0107	0.392	0.443
1p21	rs1948368,G	Imputed	5.6E-06	0.872	0.464	0.492	0.0002	0.456	0.497	0.0528	0.466	0.491	0.1550	0.488	0.493
9q31	rs7042161,T	Imputed	5.8E-06	0.867	0.325	0.349	4.E-05	0.314	0.351	0.2190	0.333	0.343	0.0091	0.327	0.348
11q24	rs544368,T	Genotyped	6.0E-06	1.221	0.139	0.120	4.E-05	0.143	0.113	0.1000	0.135	0.122	0.0590	0.139	0.131
15q25	rs2278702,T	Genotyped	6.3E-06	0.828	0.140	0.166	1.E-05	0.132	0.165	0.0089	0.139	0.161	0.9906	0.154	0.175
18p11	rs7226677,G	Imputed	6.7E-06	1.240	0.136	0.120	0.0004	0.140	0.117	0.0509	0.138	0.125	0.0367	0.129	0.126
3p26	rs1601875,A	Genotyped	6.9E-06	0.874	0.470	0.498	0.0144	0.467	0.493	0.0002	0.463	0.509	0.1834	0.483	0.494
10q22	rs703965,T	Imputed	7.5E-06	0.867	0.434	0.461	0.0065	0.446	0.467	0.2180	0.434	0.446	0.0012	0.429	0.474

The most associated SNP is shown for each region.

Supplementary Table 6: Two- or three-SNP haplotypes with strongest association in the 10q21 region.

Haplotype	SNPs in haplotype			Allele	*P*	OR	Allele frequency		r^2 with rs10994336
	SNP 1	SNP 2	SNP 3				Cases	Controls	
1	rs4582919	rs10994357	rs7910492	CCT	2.0E-09	1.41	0.083	0.065	0.803
2	rs4582919	rs7910492	rs10821773	CTC	2.2E-09	1.41	0.084	0.066	0.803
3	rs4582919	rs10994357	-	CC	2.2E-09	1.40	0.084	0.069	0.775
4	rs4582919	rs10994357	rs10821773	CCC	3.5E-09	1.41	0.081	0.064	0.817
5	rs4582919	rs10821773	-	CC	4.6E-09	1.40	0.083	0.067	0.793

Supplementary Table 7: Genotypic analysis for the 10q21, 12p13 and 15q14 regions identified in the combined analysis of the WTCCC, STEP-UCL and ED-DUB-STEP2 studies. Results are shown for the most associated genotyped SNP in each region.

		Combined	**WTCCC**	**STEP-UCL**	**ED-DUB-STEP2**
		Chromosome 10q21, SNP rs1938526			
Genotype counts	Cases	30/594/3762	15/241/1573	7/208/1244	8/145/945
	Controls	26/640/5541	12/314/2607	10/215/1782	4/111/1152
Genotypic test *P* (2df)		9.35E-08	0.0052	0.0017	0.0008
Dominance deviation *P* (1df)		0.8365	0.6926	0.2277	0.5548
Heterozygote OR, Homozygote OR		1.402,1.845	1.292,1.976	1.460,1.112	1.675,4.430
		Chromosome 12p13, SNP rs1006737			
Genotype counts	Cases	557/1992/1813	233/839/754	180/668/593	144/485/466
	Controls	679/2640/2853	336/1231/1366	193/858/924	150/551/563
Genotypic test *P* (2df)		3.03E-07	0.0027	0.0012	0.0381
Dominance deviation *P* (1df)		0.3161	0.1855	0.7998	0.8565
Heterozygote OR, Homozygote OR		1.219,1.356	1.224,1.257	1.231,1.456	1.191,1.476
		Chromosome 15q14, SNP rs2172835			
Genotype counts	Cases	358/1717/2255	159/722/899	114/549/791	85/446/565
	Controls	602/2638/2911	295/1227/1366	192/851/953	115/560/592
Genotypic test *P* (2df)		4.15E-06	0.0980	0.0009	0.0041
Dominance deviation *P* (1df)		0.618	0.9174	0.2755	0.5371
Heterozygote OR, Homozygote OR		0.840,0.742	0.901,0.824	0.780,0.732	0.808,0.562

***Supplementary Table 8**:* Breslow-Day test for homogeneity of odds ratios and exact test of Hardy-Weinberg (HW) equilibrium for the 10q21, 12p13 and 15q14 regions identified in the combined analysis of the WTCCC, STEP-UCL and ED-DUB-STEP2 studies. Homogeneity of odds ratios was tested between the three studies, but also between the seven panels that make up the three studies (WTCCC, STEP, UCL, ED, DUB and two STEP2 panels).

Chr	SNP	Breslow-Day *P*		HW equilibrium *P*	
		3 studies	7 panels	Cases	Controls
10q21	rs10994336	0.726	0.382	0.272	0.779
12p13	rs1006737	0.356	0.922	0.067	0.792
15q14	rs12899449	0.220	0.270	0.095	0.052

Supplementary Table 9: Association analyses between the 10q21, 12p13 and 15q14 regions with sex and three bipolar sub-phenotypes.

Phenotype	Case-only analysis				Case-control analysis			
	Comparison (Group 1 vs Group 2)	MA frequency		*P*	Comparison (Cases vs Controls)	MA frequency		*P*
		Group 1	Group 2			Cases	Controls	
Chromosome 10q21, SNP rs10994336								
Sex	Females vs Males	0.069	0.072	0.699	Females vs Female controls	0.069	0.055	0.00019
-	-	-	-	-	Males vs Male controls	0.072	0.051	1.66E-05
Bipolar subtype	BP2 vs BP1	0.060	0.072	0.115	BP2 vs Controls	0.06	0.053	0.39500
-	-	-	-	-	BP1 vs Controls	0.072	0.053	3.00E-08
Psychotic symptoms	Psychotic vs Non-psychotic	0.072	0.068	0.528	Psychotic vs Controls	0.072	0.053	5.73E-08
-	-	-	-	-	Non-psychotic vs Controls	0.068	0.053	0.000837
Age-at-onset	Over 18 vs 18 or under	0.072	0.066	0.217	Over 18 vs Controls	0.072	0.053	7.42E-07
-	-	-	-	-	18 or under vs Controls	0.066	0.053	0.00193
Chromosome 12p13, SNP rs1006737								
Sex	Females vs Males	0.356	0.357	0.840	Females vs Female controls	0.356	0.325	2.73E-05
-	-	-	-	-	Males vs Male controls	0.357	0.323	0.00071
Bipolar subtype	BP2 vs BP1	0.358	0.355	0.417	BP2 vs Controls	0.358	0.324	0.00034
-	-	-	-	-	BP1 vs Controls	0.355	0.324	8.76E-07
Psychotic symptoms	Psychotic vs Non-psychotic	0.356	0.355	0.946	Psychotic vs Controls	0.356	0.324	1.71E-05
-	-	-	-	-	Non-psychotic vs Controls	0.355	0.324	1.67E-05
Age-at-onset	Over 18 vs 18 or under	0.356	0.358	0.751	Over 18 vs Controls	0.356	0.324	1.97E-05
-	-	-	-	-	18 or under vs Controls	0.358	0.324	2.21E-05
Chromosome 15q14, SNP rs12899449								
Sex	Females vs Males	0.244	0.249	0.485	Females vs Female controls	0.244	0.280	5.26E-06
-	-	-	-	-	Males vs Male controls	0.249	0.271	0.01150
Bipolar subtype	BP2 vs BP1	0.249	0.245	0.867	BP2 vs Controls	0.249	0.276	0.02000
-	-	-	-	-	BP1 vs Controls	0.245	0.276	6.36E-06
Psychotic symptoms	Psychotic vs Non-psychotic	0.252	0.232	0.021	Psychotic vs Controls	0.252	0.276	0.00072
-	-	-	-	-	Non-psychotic vs Controls	0.232	0.276	1.56E-08
Age-at-onset	Over 18 vs 18 or under	0.248	0.242	0.747	Over 18 vs Controls	0.248	0.276	0.00019
-	-	-	-	-	18 or under vs Controls	0.242	0.276	1.42E-05

Supplementary Table 10: Regions of most significant interaction ($P < 5\times10^{-10}$) identified in the whole-genome epistasis analysis. The whole-genome screen was performed using our fast test for SNP-SNP interaction (*P*); results for the top interactions were also confirmed by standard logistic regression (*logit P*).

Region 1		**Region 2**		**N interactions with:**		**Top interaction**		***P***	***logit P***
Chr	Position, kb	Chr	Position, kb	$P<5\times10^{-10}$	$P<5\times10^{-8}$	SNP 1	SNP 2		
4p14	37,908	12p11	30,390	52	112	rs2055888	rs7139306	9.34E-12	1.69E-11
7p15	24,871	17q12	29,883	1	1	rs2040690	rs11656710	1.31E-11	1.18E-11
1p22	89,134	6q13	72,224	1	2	rs17471766	rs852970	9.77E-11	8.74E-10
5q13	72,726	5q31	133,898	1	1	rs4398617	rs11956902	1.56E-10	2.19E-10
10q23	87,921	16p13	5,460	1	2	rs11817682	rs583803	2.41E-10	3.36E-10
6p21	29,921	18p11	4,439	1	5	rs2523766	rs17437778	2.63E-10	1.46E-10
5q13	73,611	9p21	26,703	4	8	rs622597	rs10757623	3.35E-10	6.78E-10

4. Supplementary Note: Members of the Wellcome Trust Case Control Consortium (jun30 version)

Management Committee: Paul R Burton[1], David G Clayton[2], Lon R Cardon[3], Nick Craddock[4], Panos Deloukas[5], Audrey Duncanson[6], Dominic P Kwiatkowski[3,5], Mark I McCarthy[3,7], Willem H Ouwehand[8,9], Nilesh J Samani[10], John A Todd[2], Peter Donnelly (Chair)[11]

Data and Analysis Committee: Jeffrey C Barrett[3], Paul R Burton[1], Dan Davison[11], Peter Donnelly[11], Doug Easton[12], David M. Evans[3], Hin-Tak Leung[2], Jonathan L Marchini[11], Andrew P Morris[3], Chris CA Spencer[11], Martin D Tobin[1], Lon R Cardon (Co-chair)[3], David G Clayton (Co-chair)[2]

UK Blood Services & University of Cambridge Controls: Antony P Attwood[5,8], James P Boorman[8,9], Barbara Cant[8], Ursula Everson[13], Judith M Hussey[14], Jennifer D Jolley[8], Alexandra S Knight[8], Kerstin Koch[8], Elizabeth Meech[15], Sarah Nutland[2], Christopher V Prowse[16], Helen E Stevens[2], Niall C Taylor[8], Graham R Walters[17], Neil M Walker[2], Nicholas A Watkins[8,9], Thilo Winzer[8], John A Todd[2], Willem H Ouwehand[8,9]

1958 Birth Cohort Controls: Richard W Jones[18], Wendy L McArdle[18], Susan M Ring[18], David P Strachan[19], Marcus Pembrey[18,20]

Bipolar Disorder (Aberdeen): Gerome Breen[21], David St Clair[21]; (**Birmingham):** Sian Caesar[22], Katherine Gordon-Smith[22,23], Lisa Jones[22]; **(Cardiff):** Christine Fraser[23], Elaine K Green[23], Detelina Grozeva[23], Marian L Hamshere[23], Peter A Holmans[23], Ian R Jones[23], George Kirov[23],

Valentina Moskvina[23], Ivan Nikolov[23], Michael C O'Donovan[23], Michael J Owen[23], Nick Craddock[23]; **(London):** David A Collier[24], Amanda Elkin[24], Anne Farmer[24], Richard Williamson[24], Peter McGuffin[24]; **(Newcastle):** Allan H Young[25], I Nicol Ferrier[25]

Coronary Artery Disease (Leeds): Stephen G Ball[26], Anthony J Balmforth[26], Jennifer H Barrett[26], D Timothy Bishop[26], Mark M Iles[26], Azhar Maqbool[26], Nadira Yuldasheva[26], Alistair S Hall[26]; **(Leicester):** Peter S Braund[10], Paul R Burton[1], Richard J Dixon[10], Massimo Mangino[10], Suzanne Stevens[10], Martin D Tobin[1], John R Thompson[1], Nilesh J Samani[10]

Crohn's Disease (Cambridge): Francesca Bredin[27], Mark Tremelling[27], Miles Parkes[27]; **(Edinburgh):** Hazel Drummond[28], Charles W Lees[28], Elaine R Nimmo[28], Jack Satsangi[28]; **(London):** Sheila A Fisher[29], Alastair Forbes[30], Cathryn M Lewis[29], Clive M Onnie[29], Natalie J Prescott[29], Jeremy Sanderson[31], Christopher G Mathew[29]; **(Newcastle):** Jamie Barbour[32], M Khalid Mohiuddin[32], Catherine E Todhunter[32], John C Mansfield[32]; **(Oxford):** Tariq Ahmad[33], Fraser R Cummings[33], Derek P Jewell[33]

Hypertension (Aberdeen): John Webster[34]; **(Cambridge):** Morris J Brown[35], David G Clayton[2]; **(Evry, France): G** Mark Lathrop[36]; **(Glasgow):** John Connell[37], Anna Dominiczak[37]; **(Leicester):** Nilesh J Samani[10]; **(London):** Carolina A Braga Marcano[38], Beverley Burke[38], Richard Dobson[38], Johannie Gungadoo[38], Kate L Lee[38], Patricia B Munroe[38], Stephen J Newhouse[38], Abiodun Onipinla[38], Chris Wallace[38], Mingzhan Xue[38], Mark Caulfield[38]; **(Oxford):** Martin Farrall[39]

Rheumatoid Arthritis: Anne Barton[40], Ian N Bruce[40], Hannah Donovan[40], Steve Eyre[40], Paul D Gilbert[40], Samantha L Hider[40], Anne M Hinks[40], Sally L John[40], Catherine Potter[40], Alan J Silman[40], Deborah PM Symmons[40], Wendy Thomson[40], Jane Worthington[40]

Type 1 Diabetes: David G Clayton[2], David B Dunger[2,41], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], Barry Widmer[2,41], John A Todd[2]

Type 2 Diabetes (Exeter): Timothy M Frayling[42,43], Rachel M Freathy[42,43], Hana Lango[42,43], John R B Perry[42,43], Beverley M Shields[43], Michael N Weedon[42,43], Andrew T Hattersley[42,43]; **(London):** Graham A Hitman[44]; **(Newcastle):** Mark Walker[45]; **(Oxford):** Kate S Elliott[3,7], Christopher J Groves[7], Cecilia M Lindgren[3,7], Nigel W Rayner[3,7], Nicholas J Timpson[3,46], Eleftheria Zeggini[3,7], Mark I McCarthy[3,7]

Tuberculosis (Gambia): Melanie Newport[47], Giorgio Sirugo[47]; **(Oxford):** Emily Lyons[3], Fredrik Vannberg[3], Adrian VS Hill[3]

Ankylosing Spondylitis: Linda A Bradbury[48], Claire Farrar[49], Jennifer J Pointon[48], Paul Wordsworth[49], Matthew A Brown[48,49]

AutoImmune Thyroid Disease: Jayne A Franklyn[50], Joanne M Heward[50], Matthew J Simmonds[50], Stephen CL Gough[50]

Breast Cancer: Sheila Seal[51], Michael R Stratton[51,52], Nazneen Rahman[51]

Multiple Sclerosis: Maria Ban[53], An Goris[53], Stephen J Sawcer[53], Alastair Compston[53]

Gambian Controls (Gambia): David Conway[47], Muminatou Jallow[47], Melanie Newport[47], Giorgio Sirugo[47]; **(Oxford):** Kirk A Rockett[3], Dominic P Kwiatkowski[3,5]

DNA, Genotyping, Data QC and Informatics (Wellcome Trust Sanger Institute, Hinxton): Suzannah J Bumpstead[5], Amy Chaney[5], Kate Downes[2,5], Mohammed JR Ghori[5], Rhian Gwilliam[5], Sarah E Hunt[5], Michael Inouye[5], Andrew Keniry[5], Emma King[5], Ralph McGinnis[5], Simon Potter[5], Rathi Ravindrarajah[5], Pamela Whittaker[5], Claire Widden[5], David Withers[5], Panos Deloukas[5]; **(Cambridge):** Hin-Tak Leung[2], Sarah Nutland[2], Helen E Stevens[2], Neil M Walker[2], John A Todd[2]

Statistics (Cambridge): Doug Easton[12], David G Clayton[2]; **(Leicester):** Paul R Burton[1], Martin D Tobin[1]; **(Oxford):** Jeffrey C Barrett[3], David M Evans[3], Andrew P Morris[3], Lon R Cardon[3]; **(Oxford):** Niall J Cardin[11], Dan Davison[11], Teresa Ferreira[11], Joanne Pereira-Gale[11], Ingeleif B Hallgrimsdóttir[11], Bryan N Howie[11], Jonathan L Marchini[11], Chris CA Spencer[11], Zhan Su[11], Yik Ying Teo[3,11], Damjan Vukcevic[11], Peter Donnelly[11]

PIs: David Bentley[5,54], Matthew A Brown[48,49], Lon R Cardon[3], Mark Caulfield[38], David G Clayton[2], Alistair Compston[53], Nick Craddock[23], Panos Deloukas[5], Peter Donnelly[11], Martin Farrall[39], Stephen CL Gough[50], Alistair S Hall[26], Andrew T Hattersley[42,43], Adrian VS Hill[3], Dominic P Kwiatkowski[3,5], Christopher G Mathew[29], Mark I McCarthy[3,7], Willem H Ouwehand[8,9], Miles Parkes[27], Marcus Pembrey[18,20], Nazneen Rahman[51], Nilesh J Samani[10], Michael R Stratton[51,52], John A Todd[2], Jane Worthington[40]

[1] Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Adrian Building, University Road, Leicester, LE1 7RH, UK; [2] Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Cambridge, CB2 0XY, UK; [3] Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK; [4] Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK; [5] The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; [6] The Wellcome Trust, Gibbs Building, 215 Euston Road, London NW1 2BE, UK; [7] Oxford Centre for Diabetes, Endocrinology and Medicine, University of Oxford, Churchill Hospital, Oxford, OX3 7LJ, UK; [8] Department of Haematology, University of Cambridge, Long Road, Cambridge, CB2 2PT, UK; [9] National Health Service Blood and Transplant, Cambridge Centre, Long Road, Cambridge, CB2 2PT, UK; [10] Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Groby Road, Leicester, LE3 9QP, UK; [11] Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK; [12] Cancer Research UK Genetic Epidemiology Unit, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK; [13] National Health Service Blood and Transplant, Sheffield Centre, Longley Lane, Sheffield S5 7JN, UK; [14] National Health Service Blood and Transplant, Brentwood Centre, Crescent Drive, Brentwood, CM15 8DP, UK; [15] The Welsh Blood Service, Ely Valley Road, Talbot Green, Pontyclun, CF72 9WB, UK; [16] The Scottish National Blood Transfusion Service, Ellen's Glen Road, Edinburgh, EH17 7QT, UK; [17] National Health Service Blood and Transplant, Southampton Centre, Coxford Road, Southampton, SO16 5AF, UK; [18] Avon Longitudinal Study of Parents and Children, University of Bristol, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK; [19] Division of Community Health Services, St George's University of London, Cranmer Terrace, London SW17 0RE, UK; [20] Institute of Child Health, University College London, 30 Guilford St, London WC1N 1EH, UK; [21] University of Aberdeen, Institute of Medical Sciences, Foresterhill, Aberdeen, AB25 2ZD, UK; [22] Department of Psychiatry, Division of Neuroscience, Birmingham University, Birmingham, B15 2QZ, UK; [23] Department of Psychological Medicine, Henry Wellcome Building, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK; [24] SGDP, The Institute of Psychiatry, King's College London, De Crespigny Park Denmark Hill London SE5 8AF, UK; [25] School of Neurology, Neurobiology and Psychiatry, Royal Victoria Infirmary, Queen Victoria Road, Newcastle upon Tyne, NE1 4LP, UK; [26] LIGHT and LIMM Research Institutes, Faculty of Medicine and Health, University of Leeds, Leeds, LS1 3EX, UK; [27] IBD Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, CB2 2QQ, UK; [28] Gastrointestinal Unit, School of Molecular and Clinical Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU UK; [29] Department of Medical & Molecular Genetics, King's College London School of Medicine, 8th Floor Guy's Tower, Guy's Hospital, London, SE1 9RT, UK; [30] Institute for Digestive Diseases, University College London Hospitals Trust, London, NW1 2BU, UK; [31] Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London, SE1 7EH, UK; [32] Department of Gastroenterology & Hepatology, University of Newcastle upon Tyne, Royal Victoria Infirmary, Newcastle upon Tyne, NE1 4LP, UK; [33] Gastroenterology Unit, Radcliffe Infirmary, University of Oxford, Oxford, OX2 6HE, UK; [34] Medicine and Therapeutics, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, Grampian AB9 2ZB, UK; [35] Clinical Pharmacology Unit and the Diabetes and Inflammation Laboratory, University of Cambridge, Addenbrookes Hospital, Hills Road, Cambridge CB2 2QQ, UK; [36] Centre National de Genotypage, 2, Rue Gaston Cremieux, Evry, Paris 91057.; [37] BHF Glasgow Cardiovascular Research Centre, University of Glasgow, 126

University Place, Glasgow, G12 8TA, UK; [38] Clinical Pharmacology and Barts and The London Genome Centre, William Harvey Research Institute, Barts and The London, Queen Mary's School of Medicine, Charterhouse Square, London EC1M 6BQ, UK; [39] Cardiovascular Medicine, University of Oxford, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK; [40]arc Epidemiology Research Unit, University of Manchester, Stopford Building, Oxford Rd, Manchester, M13 9PT, UK; [41] Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 2QQ, UK; [42] Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, Magdalen Road, Exeter EX1 2LU UK; [43] Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula Medical School, Barrack Road, Exeter EX2 5DU UK; [44] Centre for Diabetes and Metabolic Medicine, Barts and The London, Royal London Hospital, Whitechapel, London, E1 1BB UK; [45] Diabetes Research Group, School of Clinical Medical Sciences, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK; [46] The MRC Centre for Causal Analyses in Translational Epidemiology, Bristol University, Canynge Hall, Whiteladies Rd, Bristol BS2 8PR, UK; [47] MRC Laboratories, Fajara, The Gambia; [48] Diamantina Institute for Cancer, Immunology and Metabolic Medicine, Princess Alexandra Hospital, University of Queensland, Woolloongabba, Qld 4102, Australia; [49] Botnar Research Centre, University of Oxford, Headington, Oxford OX3 7BN, UK; [50] Department of Medicine, Division of Medical Sciences, Institute of Biomedical Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; [51] Section of Cancer Genetics, Institute of Cancer Research, 15 Cotswold Road, Sutton, SM2 5NG, UK; [52] Cancer Genome Project, The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; [53] Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK; [54] PRESENT ADDRESS: Illumina Cambridge, Chesterford Research Park, Little Chesterford, Nr Saffron Walden, Essex, CB10 1XL, UK.

References

1. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
2. Sklar P, S.J., Fan J, Ferreira MAR, Perlis RH, Chambert K, Nimgaonkar VL, et al. Whole-genome Association Study of Bipolar Disorder. *Mol Psychiatry* (in press).
3. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
4. Rice, J. et al. The familial transmission of bipolar illness. *Arch Gen Psychiatry* **44**, 441-7 (1987).
5. Association, A.P. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*, (American Psychiatric Association, Washington, DC, 1994).
6. Cardno, A.G. et al. Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry* **56**, 162-8 (1999).
7. Joyce, P.R. et al. Affective disorders in the first-degree relatives of bipolar probands: results from the South Island Bipolar Study. *Compr Psychiatry* **45**, 168-74 (2004).
8. Kieseppa, T., Partonen, T., Haukka, J., Kaprio, J. & Lonnqvist, J. High concordance of bipolar I disorder in a nationwide sample of twins. *Am J Psychiatry* **161**, 1814-21 (2004).
9. Simpson, S.G. et al. Bipolar II: the most common bipolar phenotype? *Am J Psychiatry* **150**, 901-3 (1993).
10. Edvardsen, J. et al. Heritability of bipolar spectrum disorders. Unity or heterogeneity? *J Affect Disord* **106**, 229-40 (2008).
11. Li Y & GR, A. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. *Am J Hum Genet* **S79**, 2290 (2006).
12. Reich, D.E. & Goldstein, D.B. Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* **20**, 4-16 (2001).
13. Freedman, M.L. et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet* **36**, 388-93 (2004).
14. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
15. Perlis, R.H. et al. Predictors of recurrence in bipolar disorder: primary outcomes from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *Am J Psychiatry* **163**, 217-24 (2006).